1ST WORKSHOP ON

Safeguarding National Infrastructures

EDITOR: CHRIS JOHNSON

GIST TECHNICAL REPORT G2005-2,
DEPARTMENT OF COMPUTING SCIENCE, UNIVERSITY OF GLASGOW, SCOTLAND.

ACKNOWLEDGEMENTS

The workshop has also been supported by the EC ADVISES Research Training Network.

We also acknowledge the support of the following organisations:

Glasgow Accident Analysis Group

UNIVERSITY of GLASGOW

British HCI Group
www.bcs.org.uk/hci

TABLE OF CONTENTS

THURSDAY 25TH AUGUST

09.00-09.30 C. Johnson	Welcome and Introduction.
09.30-10.30 Chair: Chris Johnson (Univ. of Glasgow)	**Invited Talk 1:** *Protecting Infrastructures: The Role of Risk Analysis.* George Apostolakis, MIT, USA.
10.30-11.00	*Coffee*
11.00-12.30 Chair: Nick Chozos (Univ. of Glasgow)	**Paper Session 1: Threat Assessment in Computer Network Defence** *Security Threat Assessment across Large Network Infrastructures* Grigorios Fragkos and Andrew Blyth, School of Computing, University of Glamorgan, UK. *Exploring the Role of Deception in Computer Network Defence and Information Operations Environment* Zafar Kazmi, Theodore Tryfonas and Stilianos Vidalis, School of Computing, University of Glamorgan, UK
12.30-14.30	*Lunch*
14:30-16:00 Chair: Thomasz Mistrzyk (Univ. of Paderborn/Univ of Glasgow)	**Paper Session 2: Safeguarding Transportation Infrastructures** *Denial of Safety Critical Services of a Public Mobile Network for a Critical Transport Infrastructure* Ester Ciancamerla and Michele Minichino, ENEA CR Casaccia, Rome, Italy. *EUROCONTROL's Systemic Occurence Analysis Methodology (SOAM)* *Tony Licu, EUROCONTROL, Brussels, Belgium.* Brent Hayward and Andrew Lowe, Dedale, Melbourne, Australia.
16:00-16:15	*Tea*
16:15-17:00 Chair: Bastiaan A. Schupp, (Univ of York, EC Research Center, Ispra)	**Paper Session 3: Emerging Failures** *Safeguarding Information Intensive Critical Infrastructures Against Novel Types of Emerging Failures* C. Balducelli, S. Bologna, L. Lavalle and G. Vicoli, ENEA, Italian Agency for New Technology, Energy and the Environment, Rome.
18:00-20:00	**Informal Reception**

FRIDAY 26TH AUGUST

09.30-10.00 C. Johnson	*Welcome and Coffee*,
10.00-10.30 Chair: Richard Sinnott, (Univ. of Glasgow)	**Invited Talk 2:** *Critical Information Infrastructure Protection: The Role of the UK National Infrastructure Security Coordination Center* Mike Corcoran, UK National Infrastructure Security Coordination Center.
10.30-11.00	*Coffee*
11.30-12.30 Chair: Karen Renaud (Univ. of Glasgow)	**Paper Session 4: Risk and Interdiction** *Methodology for Identifying Near-Optimal Interdiction Strategies for a Power Transmission System* Vicki Bier, Eli Gratz, Naraphorn Haphuriwat, Wairimu Magua, Kevin Wierzbicki University of Wisconsin-Madison, Madison, USA. *Identification of Critical Locations across Multiple Infrastructures for Terrorist Actions* Sean Patterson and George Apostolakis, Massachusetts Institute of Technology, USA
12.30-14.30	*Lunch*
14:30-16:00 Chair: Phil Gray (Univ. of Glasgow)	**Paper Session 6: Security and Privacy** *Passwords, Passwords Everywhere and Not a Minute to Think* Karen Renaud University of Glasgow, UK. *Prospects for a Robust Electronic Voting System for the UK* Ishbel Duncan and Tim Storer, University of St. Andrews, UK.
16.00-16.15	*Tea*
16.15-17:00	**Paper Session 6**: Closing Paper (Time permitting) *A Risk-Based Approach to the Evacuation of Large Public Buildings: Hospitals* Or *Root Cause Analysis for Complex Security Incidents* Chris Johnson University of Glasgow, UK
17:00-17:15	*Close and hand-over.*

SATURDAY, 27TH AUGUST

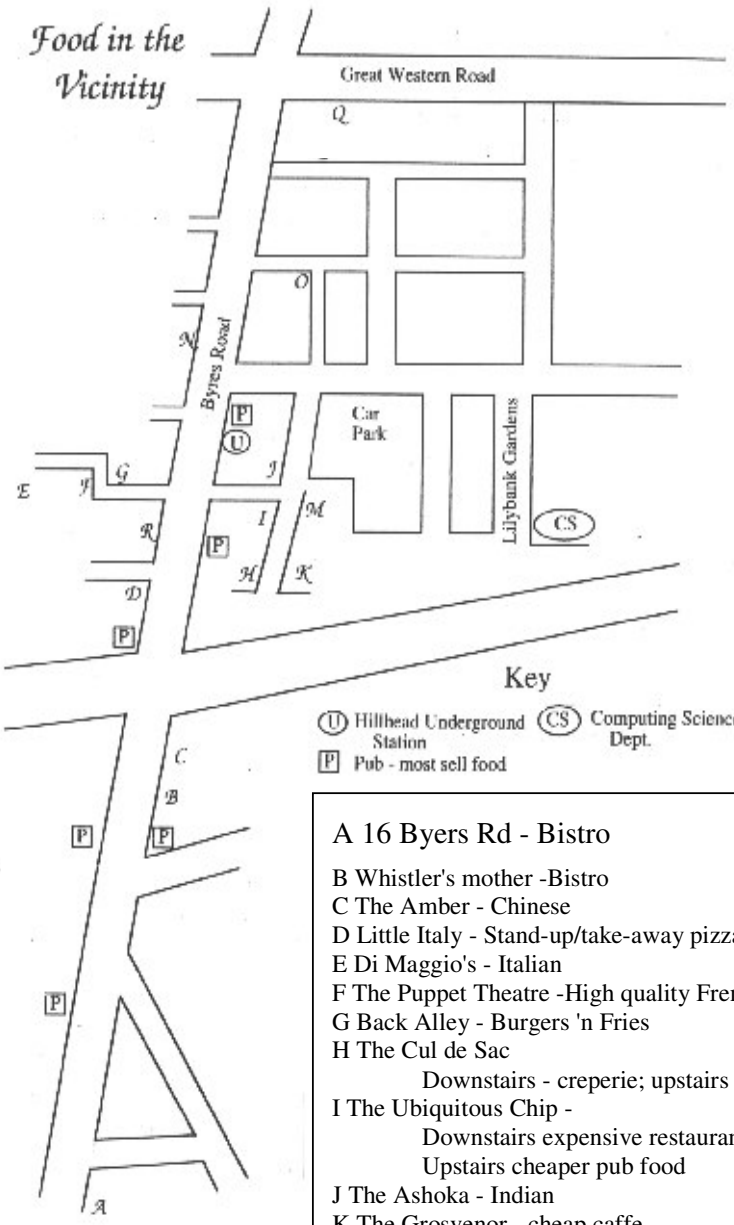
This will provide the opportunity for informal discussions about the issues raised during the workshop. The day will be spent on the Isle of Arran, off the west Coast of Scotland. The intention is to meet outside the Department of Computer Science at 07:30. We will be taking the train because this connects directly with the CalMac (http://www.calmac.co.uk) ferry onto the Island. Anyone who misses the first rendez-vous can meet us underneath the large clock at Central Station for 08:00 (Buchanan Street is the nearest Underground station). Trains depart from Glasgow Central station at 08:33, arrives at Ardrossan harbour at 09:25. The ferry leaves for Arran at 09:45. Ferry arrives at Brodick on Arran at 10:40. The ferry departs Brodick at 16:40, arrives Ardrossan 17:35. The train arrives at Glasgow Central 18:52. There is an additional service departing Brodick at 19:20, arriving at Ardrossan to connect with the 20:30 that arrives into Glasgow at 21:22.

If anyone misses this departure then they will have to spend the night on the Island (there are lots of hotels and bed & breakfast places). Arran Tourist Office can be contacted on 01770-302140 or 01292 678100 (http://www.ayrshire-arran.com/arran.htm) for hotel accommodation and other enquiries. The whiskey distillery is open for visits from 10.00-18.00 and can be contacted on 01292 678100.

Out	Monday to Saturday					Sunday			
Glasgow Central dep	**0833**	1115	1415	1650	1915 Friday only	0840	1115	1405	1655
Ardrossan dep	**0945**	1230	1515	1800	2030	0945	1230	1515	1800
Brodick arr.	**1040**	1325	1610	1855	2125	1040	1325	1610	1855

Return	Monday to Saturday						Sunday			
Brodick dep	0820	1105	1350	**1640**	1920	2140 Fridays only	1105	1350	1640	1920
Ardrossan arr	0915	1200	1445	**1735**	2015	2235	1200	1445	1735	2015
Glasgow Central arr	1022	1322	1622	**1852**	2122	-	1328	1550	1850	2117

Restaurants in the Local Area



A 16 Byers Rd - Bistro

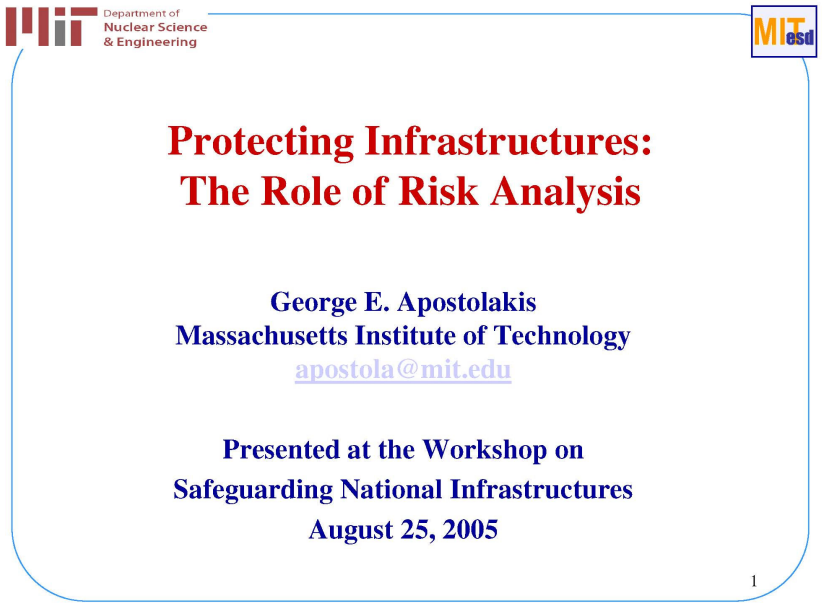
B Whistler's mother -Bistro
C The Amber - Chinese
D Little Italy - Stand-up/take-away pizzas
E Di Maggio's - Italian
F The Puppet Theatre -High quality French
G Back Alley - Burgers 'n Fries
H The Cul de Sac
Downstairs - creperie; upstairs - pub
I The Ubiquitous Chip -
Downstairs expensive restaurant
Upstairs cheaper pub food
J The Ashoka - Indian
K The Grosvenor - cheap caffe
M Jinty McGinty's - Irish pub & bar food
O The Metro - salads, hot dishes
Q The Parthenon - Greek food
R Burger King

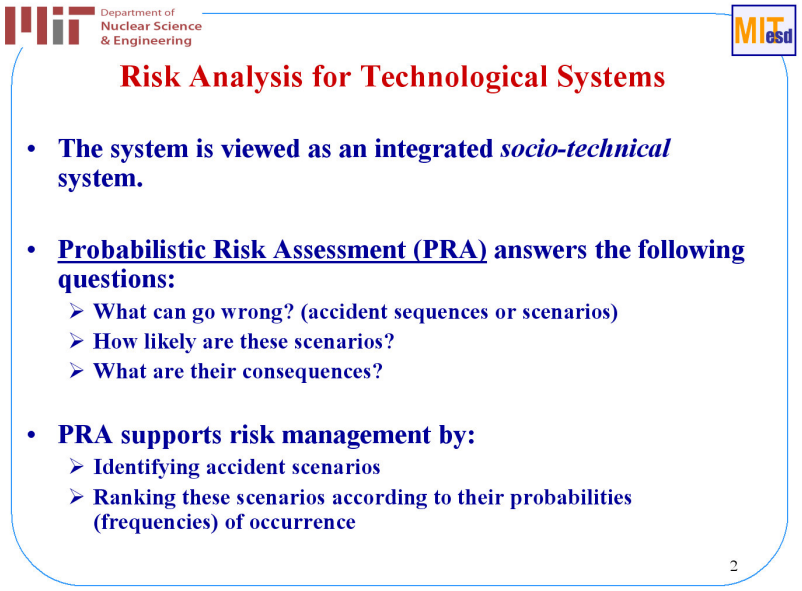
Protecting Infrastructures: The Role of Risk Analysis.

George Apostolakis,

MIT Department of Nuclear Science and Engineering
77 Massachusetts Avenue, 24-221
Cambridge, MA 02139-4307
MIT, USA

Email: apostola@mit.edu





Department of Nuclear Science & Engineering

MIT esd

Observations on Infrastructures

- Large, diffuse, inter-connected networks, as opposed to well defined systems such as nuclear power plants and the International Space Station.
- Difficult to analyze using top-down conventional mathematical theories, such as Probabilistic Risk Assessment.
- Infrastructure systems were never intended by their designers to resist the consequences of planned malicious destruction.

3

Department of Nuclear Science & Engineering

MIT esd

Objective

- To rank the elements of an infrastructure according to their risk value for "random" failures or their vulnerability to terrorism.
- In both cases, the value of the element to the Decision Maker (DM) is assessed using a value tree and disutility functions.
- For random failures, the expected disutilities are the basis for ranking the infrastructure elements. This decision rule is borrowed from Multiattribute Utility Theory, the only normative decision-making theory.
- For malevolent acts, probabilities are difficult to evaluate. The element's value is combined with its susceptibility to attack to develop a vulnerability ranking.

4

Department of Nuclear Science & Engineering

MIT esd

The Case Studies

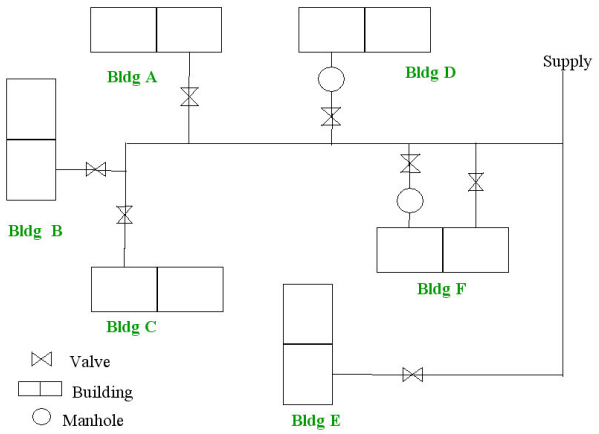
- **Specific Assets**
  - Six buildings on the MIT campus
  - Three infrastructures (electric power, water, natural gas)
  - Binary logic for the elements
  - Critical locations identified via minimal cut sets
- **A Town**
  - Water infrastructure of a European city
  - Network's capacity and time included.
- **In both cases**
  - An objectives hierarchy (value tree) is developed with the DM
  - The threat is assumed to be minor.

5

Department of Nuclear Science & Engineering

MIT esd

MIT Natural Gas Infrastructure

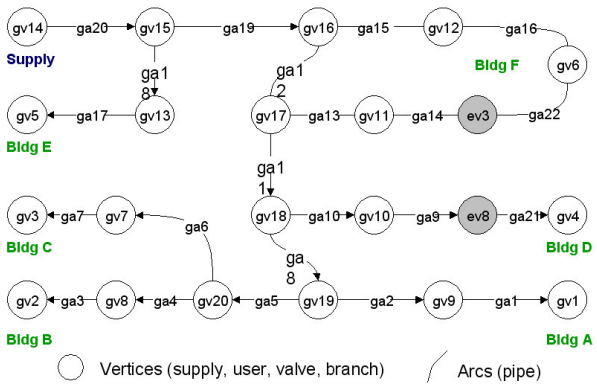


6

Department of Nuclear Science & Engineering

MIT esd

Natural Gas Network Digraph



7

Department of Nuclear Science & Engineering

MIT esd

Minimal Cut Sets (Scenarios) Affecting the Assets

Number of mcs	Number of Users Impacted
7	6
7	5
11	4
107	3
256	2
275	1

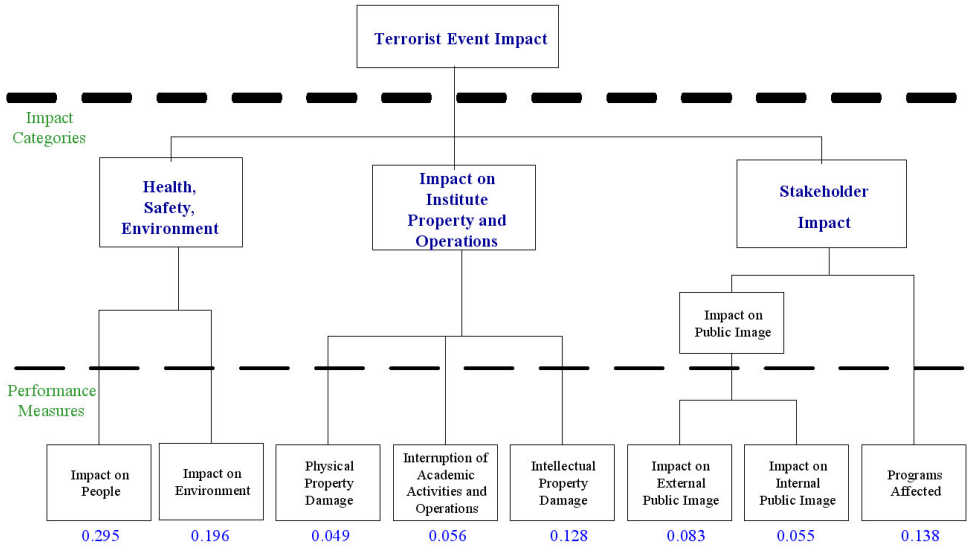
663 total mcs

Example: (ev1, ev2) impacts electrical service to all six buildings.

8

Department of Nuclear Science & Engineering

The MIT Value Tree



9

Department of Nuclear Science & Engineering

Constructed Scale:
Interruption of Academic Activities & Operations

Level	Description	Disutility
4	Extreme interruption (greater than 6 months)	1.00
3	Major interruption (1 to 6 months)	0.57
2	Moderate interruption (1 to 4 weeks)	0.19
1	Minor interruption (less than 1 week)	0.06
0	No interruption	0.00

10

Department of Nuclear Science & Engineering

MIT esd

Prioritization Methodology

- **Performance Index (expected disutility)**

$$\overline{PI}_j = \sum_i^{K_{pm}} w_i \bar{d}_{ij}$$

$\overline{PI}_j$ expected performance index for vulnerability *j*

w_i weight of the performance measure *i*

$\bar{d}_{ij}$ expected disutility of performance measure *i* for vulnerability *j*

K_{pm} number of performance measures

11

Department of Nuclear Science & Engineering

MIT esd

Problems with Expected Values

$$\bar{d}_{ij} = Pr(A \cap mcs_j) x \sum_k \left[Pr(d_i^k / mcs_j) x d_i^k \right]$$

$$\mathbf{Pr(A \cap mcs_j) = Pr(A) x Pr(mcs_j / A)}$$

- **Assume the presence of a minor threat → Ignore Pr(A).**
- **Handle $Pr(mcs_j/A)$ qualitatively in susceptibility assessment.**
- **$Pr(d_i^k / mcs_j)$ Probability of disutility level k for PM_i**
- **Assess disutility level k for PM i conservatively**

12

Department of Nuclear Science & Engineering

MIT esd

Performance Index for Screening

$$PI_j = \sum_{i=1}^{K_{pm}} w_i d_{ij}$$

PI_j performance index for minimal cut set j

d_{ij} disutility of performance measure *i* for minimal cut set *j* (assessed conservatively)

13

Department of Nuclear Science & Engineering

MIT esd

PI Values for MCS

PI	Number of mcs	Minimal cut sets
0.24742	1	(ev1,ev2)
0.15881	47	(ev23,ev6), (ev1,ev5), (ea20,ev4), …….
0.11508	1	(ev8)
0.11370	3	(wv14), (wv15), (wa20)
0.09391	48	(ev21,ev6), (ev20,ev5), (ea17,ev2), …….
0.09030	2	(wv16), (wa19)
0.08861	55	(ev24,ev42), (ea40,ev42), (ea39,ev38), …….
.....	.....	.....

14

Department of Nuclear Science & Engineering

MIT esd

Susceptibility Categories

- Extreme - Completely open, no controls, no barriers
- High - Unlocked, non-complex barriers (door or access panel)
- Moderate - Complex barrier, security patrols, video monitor
- Low - Secure area, locked, complex closure
- Very Low - Guarded, secure area, locked, alarmed, complex closure
- Zero - Completely secure, no susceptibility

15

Department of Nuclear Science & Engineering

MIT esd

Vulnerability Ranking

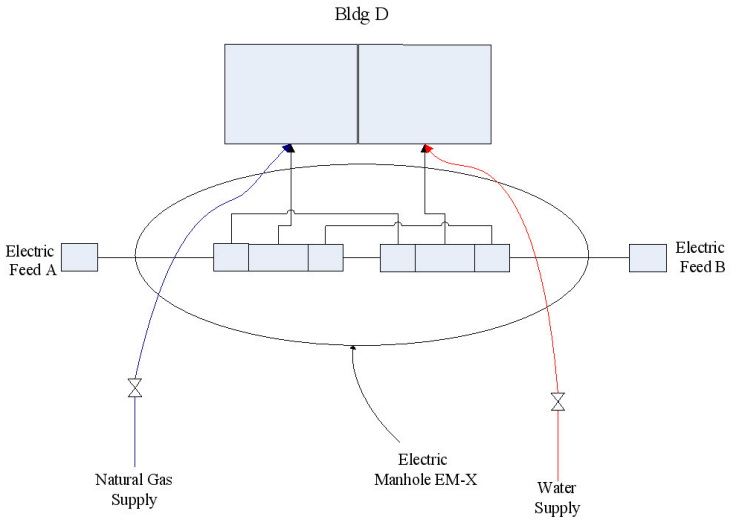
Vulnerability Category	Number of mcs	Minimal Cut Sets
I (Red) A severe vulnerability in the infrastructure requiring the most immediate attention.	1	(ev8)
II (Orange)	0	none
III (Yellow)	5	(ev21), (ev22), (ev3), (ev34), (ev9)
IV (Blue)	19	(wa20), (wv14), (wv15), (ev11), (ev18), (ev19), (ev25), (gv1), (gv2), (gv3), (gv4), (gv5), (gv6), (wv1), (wv2), (wv3), (wv4), (wv5), (wv6)
V (Green) Low (and below) on the susceptibility scale and low (and below) on the value scale.	638	All remaining mcs

16

Department of Nuclear Science & Engineering

MIT esd

Red MCS: *ev8* (Electric Manhole, PI = 0.11508)



17

Department of Nuclear Science & Engineering

MIT esd

Input to Risk Management

Category	Possible Countermeasures
Reduce Susceptibility	•Weld the manhole cover (Yellow) •Alarm the manhole cover •Monitor the manhole cover •Increase the security patrols
Reduce Value	•Install additional independent infrastructure supply lines (Yellow) •Install internal (to the building) back-up sources

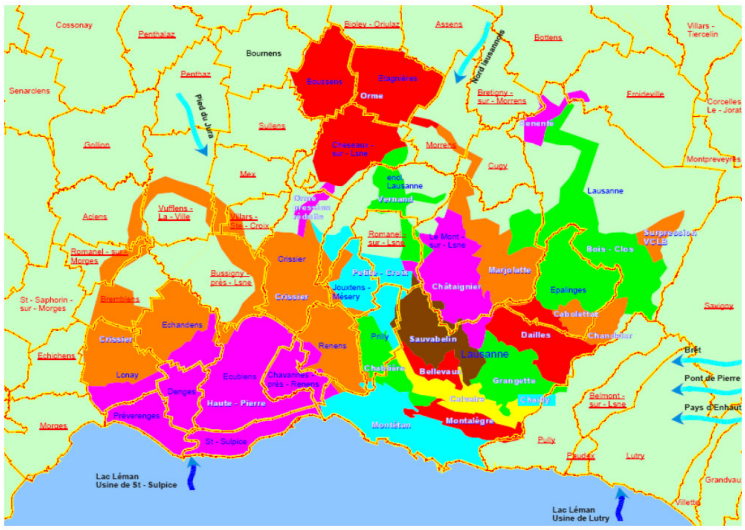
1. The PI for building D, for all three services, is 0.11508. The electric service accounts for 63% of that value. Installing an additional electrical feed to the building would reduce the PI of the manhole to 0.04234, resulting in "yellow." Expensive fix.
2. If the additional electrical service is installed and the manhole cover is welded, the new vulnerability category for manhole EM-X (mcs *ev8*) would be assessed to be green.

18

Department of Nuclear Science & Engineering

MIT esd

City Water Service (Pressure Zones)



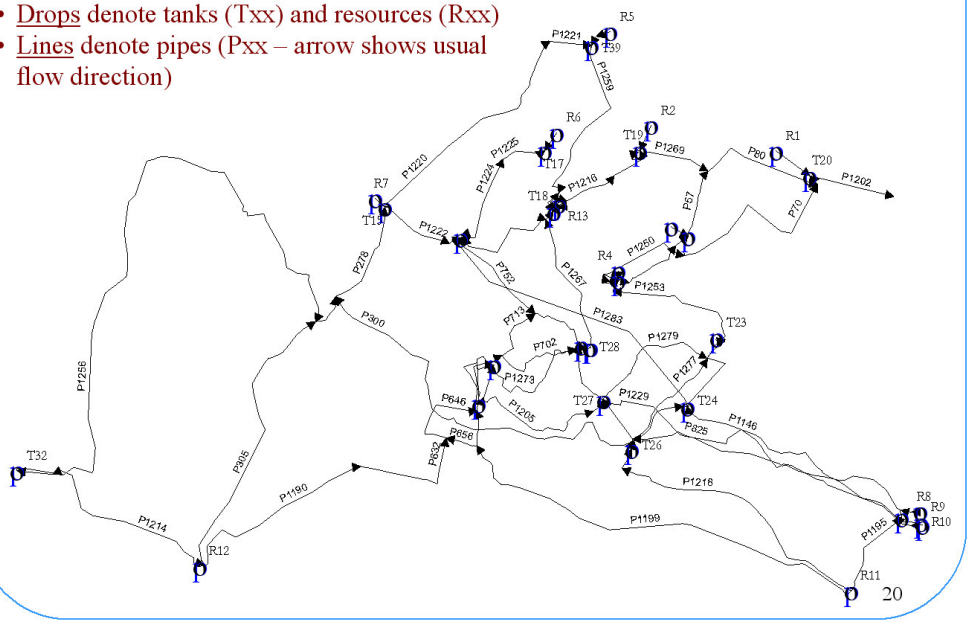
19

Department of Nuclear Science & Engineering

MIT esd

Water Supply Network

- Drops denote tanks (Txx) and resources (Rxx)
- Lines denote pipes (Pxx – arrow shows usual flow direction)

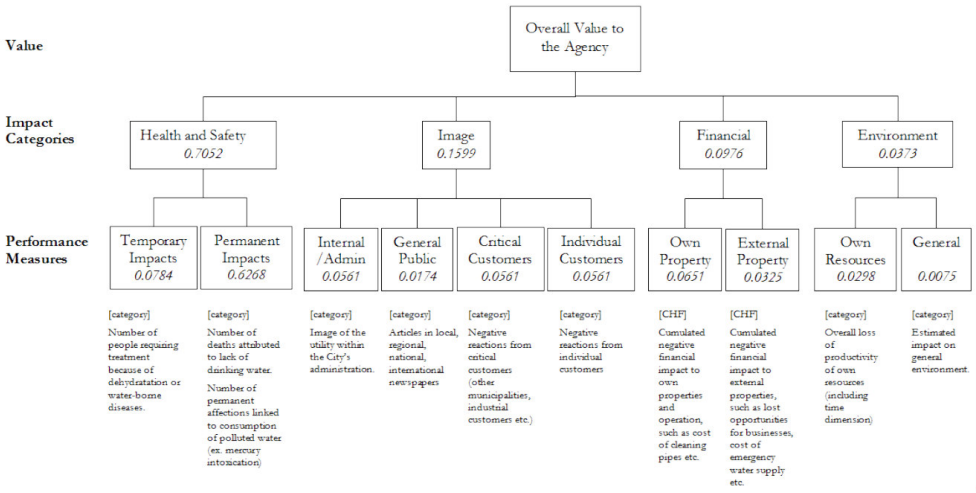


20

Department of Nuclear Science & Engineering

MIT esd

Value Tree of the Water Supply Agency



21

Department of Nuclear Science & Engineering

MIT esd

Consequence Matrices

Objective	Level	Description	Individual Customers		
		Unit	Number		
		Duration of Interruption	6 hrs	1 day	1 week
Temporary Impacts	4	A large share of the served population requires treatment because of water-borne conditions	NA	NA	NA
	3	Hundreds of persons require treatment, dozens of them hospitalization for water-borne conditions	NA	NA	100,000
	2	Dozens of persons require treatment, some of them hospitalization for water-born conditions	100,000	100,000	50,000
	1	A few persons require light treatment for water-borne conditions	50,000	50,000	20,000
	0	No health impact	0	0	0

22

Department of Nuclear Science & Engineering

MIT esd

Risk Categories (Random Failures)

Scenarios for which (frequency of occurrence) × (value) is greater than or equal to…	… belong to category…	# of elements in this category		
		Pipes	Sources & Tanks	Total
0.32231406	1	0	0	0
0.09917356	2	0	2	2
0.02479340	3	0	1	1
0.00000001	4	28	14	42
0.00000000	5	114	16	130
Total		142	33	175

23

Department of Nuclear Science & Engineering

MIT esd

Risk Prioritization Map (Random Failures)

24

Department of Nuclear Science & Engineering

MIT esd

Vulnerability Prioritization Map (Malevolent Acts)

25

Department of Nuclear Science & Engineering

MIT esd

Top Scenarios and Vulnerabilities of the System

Element	Value	Random failure category	Vulnerability category due to a malevolent act
R11	0.237457	2	II
R12	0.278468	2	II
T8	0.009228	3	IV
T27	0.040265	4	III
P916	0.036691	4	IV
P1239	0.035634	4	III

26

Department of Nuclear Science & Engineering

MIT esd

Risk Management

- **The utility had been considering an increase in the transport capacity from the right side of the system to the left by creating a supplementary connection from tank T24 to tank T16. This connection would serve as a redundant supply of water to the left side of the network in case resource R12 became unavailable.**
- **In order to evaluate this proposed addition, we recomputed all the single-element scenarios to assess their updated categories. The impact of the proposed increase in transport capacity was minimal (a few shifts from vulnerability category IV to V in the malevolent action threat analysis). As a result of our analysis, the utility is now reconsidering the effectiveness of this particular proposal.**

27

Department of Nuclear Science & Engineering

MIT esd

Bibliography

- Amin, M. (2002). "Toward Secure and Resilient Interdependent Infrastructures." Journal of Infrastructure Systems **8**(3): 67-75.
- Apostolakis, G. E. (2004). "How Useful is Quantitative Risk Assessment?" Risk Analysis **24**(3): 515-520.
- Apostolakis, G. E. and D. M. Lemon (2005). "A Screening Methodology for the Identification and Ranking of Infrastructure Vulnerabilities Due to Terrorism." Risk Analysis **25**(2): 361-376.
- Garrick, B. J., J. E. Hall, M. Kilger, J. McDonald, J. McGroddy, T. O'Toole, P. Probst, E. Rindskopf Parker, R. Rosenthal, A. Trivelpiece, L. Van Arsdale, and E. Zebroski (2004). "Confronting the risks of terrorism: making the right decisions." Reliability Engineering and System Safety **86**(2): 129-176.
- Haimes, Y. Y. and B. M. Horowitz (2004). "Modeling Interdependent Infrastructures for Sustainable Counterterrorism." Journal of Infrastructure Systems **10**(2): 33-42.
- Karydas, D.M. and Gifun, J.F. (2005) "A Method for the Efficient Prioritization of Infrastructure Renewal Projects in the Massachusetts Institute of Technology Department of Facilities," Reliability Engineering and System Safety, in press.
- Michaud, D. and Apostolakis, G.E. (2005). "Screening Vulnerabilities in Water-Supply Networks," Submitted for publication to the Journal of Infrastructure Systems.
- Paté-Cornell, M.E. and Guikema, S. (2002). "Probabilistic Modeling of Terrorist Threats: A Systems Analysis Approach to Setting Priorities Among Countermeasures," Military Operations Research, **7**:5-20.

28

Security Threat Assessment across Large Network Infrastructures

Grigorios Fragkos, Andrew Blyth

School of Computing, University of Glamorgan, Cardiff, CF37 1DL, Wales
{gfragkos, ajcblyth}@glam.ac.uk

Abstract: Despite the advantages by the Intrusion Detection community and Computer Network Defense, network infrastructures still suffers from the danger of targeted and untargeted network attacks. Most of the ongoing research is focused on protecting a single network or even a larger infrastructure without providing the bigger picture of how to protect a number of large homogeneous and heterogeneous network infrastructures. This can be achieved by combining their existing capabilities and by making them to work together in order to develop a holistic picture of how to perform network defense. Also, the need for more dynamic solution in the area of threat assessment by following the path of real-time analysis is presented. Finally, this work tries to explain, in realistic terms, up to what level security can be considered achievable and how existing intelligent technologies could be used in order to reach this point.

Keywords: Security, Threat Assessment, Large Network Infrastructures, Intrusion Detection Systems, Computer Network Defense, Real Time, NISCC, CNI

Introduction

Security incidents in the last years have been increasing so rapidly that the algebraic definition "geometrical progression" could be used in order to describe/characterise their exponential increment. The reported incidents have evolved from 6 in 1988 to 21756 in 2000 and consequently to 137529 in 2003 [CERT 2003]. Due to the fact that enterprises, companies, organization, large/small businesses often have serious reasons not to report security incidents and keep secret any damage that might have occurred, the alluded number in the real world, is definitely larger.

"*Many companies still seem unwilling to report e-crime for fear of damaging their reputation*," says Larry Johnson, Special Agent in Charge, Criminal Investigative Division, United States Secret Service. "*However, as we see with this survey, ignoring the problem or dealing with it quietly is not working. The question is not why can't we stop these criminal acts from happening, but rather, why are we allowing them to take place? The technology and resources are there to effectively fight this. We just need to work smarter to do it*."

Referring to the above quote made by Larry Johnson and especially where technology and resources are pointed out we will try to answer questions like: **a**) why do network infrastructures still suffer from attacks and why do we still wondering why we cannot deal efficiently with the security related issues by taking active countermeasures against them. **b**) Should today's security, still be considered as a technology problem? **c**) How and what kind of system, built with security in mind, could protect large network infrastructures efficiently by performing threat assessment?

Defining Security

Within the ISO 17799 the term "Information Security" is defined as *"Security Preservation of confidentiality, integrity and availability"* and "Threat" is defined as *"A potential cause of an icident that may result in harm to a system or organization"*. A prior requirement to distinguish before conducting any discussions considering the process of threat assessment in large infrastructures, we should try to have a complete understanding of the word "security".

The Cambridge Dictionary describes security as:
"*The ability to avoid being harmed by any risk, danger or threat*"

Also, the Oxford English Dictionary describes security as:
"*The state of being or feeling secure*"

In both cases the definitions lack to describe the word security for obvious reasons. In the first case it is impossible in any way to predict and avoid being harmed from ANY risk, ANY danger and ANY threat. In

order to make this definition more realistic we will add a single word and alter the definition to: "*The ability to avoid being "irrecoverably" harmed by any risk, danger or threat*". On the other hand, the latter definition fails to determine the level of the "..state of being or feeling secure". Consequently, which is the level of security which can be considered as secure enough? Also, Oxford's definition, by not defining the desirable level of the "state", fall into a recursive loop when including the word "secure" within the definition, where "secure" is described as "*protected against attack or other criminal activity*".

Therefore the definitions should not be considered as absolute definitions/descriptions of the word "security" in the real world due to the fact that they individually describe a practically impossible goal. Consequently, in order to describe security in a more realistic way we could combine the two definitions and define security as:

The state of being or feeling secure, by having the ability to avoid being irrecoverably harmed by any risk, danger or threat, when/for protecting a specific asset. As the word "irrecoverably" does not exit in the dictionary as an adverb of the existing word "irrecoverable" (adjective) we should rearrange the sentence in order to be exact and absolute concerning the words used within the definition. Thus, the final definition should define security as:

The state of being or feeling secure, by having the ability to avoid being harmed at an irrecoverable level, by any risk, danger or threat, when/for protecting a specific asset. (Author's definition, where "secure" is defined according to the Oxford's dictionary definition)

Applying the definition in real life and considering as an asset any key element (small or large) that we should be securing (i.e. Computer's Password, Server Room, University's Network) it is easier to identify/set independently the required state/level of security that is necessary for each element. Thus, having in mind the above definition, we will try to describe why the architecture presented later in the paper could be considered suitable to perform security threat assessment in large network infrastructures and how this architecture sets the frontiers by establishing the proper intelligent defend mechanism to protect our network.

NISCC, CNI, and Smart Procurement

The National Infrastructure Security Co-ordination Centre (NISCC) was set up in 1999 based on government resources from many departments (Defense, Central Government Policy, Trade, the Intelligence Agencies and Law Enforcement) in order *to ensure the continuity of society in time of crisis*. *"A fundamental role for any government is to ensure the continuity of society in times of crisis. This often involves providing extra protection to essential services and systems to make them more resistant to disruption and better able to recover quickly"* [NISCC 2005]. These essential services and systems mentioned above are known in the U.K. as Critical National Infrastructure (CNI) and due to the fact that NISCC cooperates with different types of infrastructures without limiting itself by geographical borders. Thus, it allows the bridge of distributed sources of information in order to be able to conduct Threat Assessment, Outreach, Response and Research & Development [NISCC 2005]. On the other hand CNI specifically deals with *assets, services and systems that support the economic, political and social life of the UK whose importance is such that any entire or partial loss or compromise could:*

- *cause large scale loss of life*
- *have a serious impact on the national economy*
- *have other grave social consequences for the community*
- *be of immediate concern to the national government*

[NISCC 2005]

Thinking of the types of systems, services and assets which the above points have direct impact upon it is obvious that we will come up with categories in various sectors. In the UK, the *CNI is categorized as ten interdependent sectors: Communications, Emergency Services, Energy, Finance, Food, Government & Public Service, Health, Public Safety, Transport and Water* [NISCC 2005]. Consequently, security incidents having as target such types of infrastructures may have catastrophic results, not only in a limited and restricted form but in a wider and broadly manner.

Having in mind the above information about the already existed NISCC we should try to take a step back and see the bigger picture and the resources that could be beneficial in terms of securing such assets. We could try to expand existed computer and network defensive technologies by combining them with the information and services provided by the NISCC. Thus, to design and develop a mechanism or some prototype architecture that could easily applied in large infrastructures. This will have as result to control and prevent

further damage by mitigating external risks. For example, having in mind the health sector and by choosing Hospitals from various geographical locations we could have an information gathering mechanism in order to assess the security risks and threats from around the world. A similar idea can be applied to the education sector and consequently to any other large infrastructure that could interchange information for a common purpose. In the following schematic approach [Figure 1] a primarily design of the alluded idea is presented.

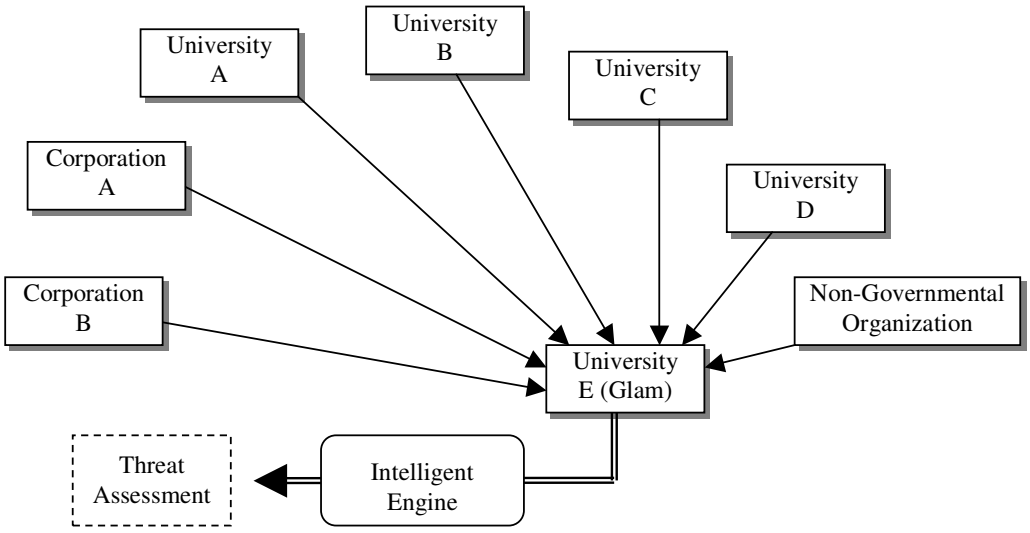


Figure 1 – Primarily design of the Security Threat Assessment

The process of threat assessment is our final goal but before getting to that point we should analyse the nature of the "Intelligent Engine". This part of the architecture is based on technologies currently available and which we will discuss later on. Until now we managed to propose and describe a future implementation of a wider approach around the idea of security threat assessment. At this point, the opportunity is given to understand why the definition of the word security has been redefined previously. According to the novel definition we are in a position to understand how important every asset is, in such critical systems which we are trying to protect, (even in these large scale environments) and thus we can define the level of security to be applied for each and every one of them.

Of course, today's Intrusion Detection Systems can help us protect these kinds of networks but as it was mentioned previously the actual merge and cross reference of information gathering generated from various sources must take place manually and by a human. Thus, in a larger or a series of larger networks the security analyst will have to spend days or even many months in order to produce a final report. Consequently, the problem security experts have when defending such networks is the lack of an intelligent engine that will minimize or even eliminate the vast and unnecessary amount of information. Moreover, this intelligent and efficient way should minimize and finally try to eliminate the amount of time spent from the moment the attack has started until the moment our defending system has picked up the ongoing attack. In other words, the need of an intelligent system that will be capable of performing security threat assessment in real-time should be considered mandatory in the future.

Despite the implementation issues surrounding such a project we should be aware of the cost and manpower needed for this type of network interconnection. However, we should also be aware of the cost and the consequences if a successful attack takes place against a large infrastructure (i.e. Electricity Distribution Centres in the U.K.). Thus, the idea of smart procurement should be taken seriously under consideration at this point of time [Humphry 1998]. Smart procurement refers to the financial issues arising when we have to deal with such large projects. In a similar way the Ministry of Defence of the U.K. is applying Smart Procurement in order to calculate if the amount of available resources needed for purchasing military equipment, is equivalent to the amount of equipment they need to purchase [MoD 2001]. Apparently, larger and larger quantities of financial resources and man-hours are spent on setting security frameworks in enterprises and organizations. However, this has never been a stopping point for attacker due to statistical proofs [Goodwin 2002].

In our case scenario Smart Procurement can be used and applied in a very efficient way due to the fact that existing network infrastructures have the required hardware in order to help us build our defending system. They are composed of computer systems that can be used as sensors for reporting all the network activity to the Intelligent Engine. The question now is, how and in what way is this feasible from the today's technological point of view, and of course if we have some or any limitations in developing such a system today.

State of the Art Network Defending Systems

The Intrusion Detection community is trying to automate the process of identifying, analyzing and responding to procedures when security incidents have been identified. Every logging-capable network component such as routers, firewalls, IDS, honey pots, etc. is generates a vast amount of audit data. The cross-reference, merge, analysis and assessment of such information must still be done by humans [Blyth 2003] in order to be efficient and capable of concluding into final results. The manual response is not feasible especially when an attack has targeted a larger scale network infrastructure and even more when the attack is against multiple network infrastructures.

The heart of Intrusion Detection Systems (IDS) relies on four groups according to the technology used to detect events; Network Based, Host Based, Application Based and Stack Based [Debar 1999], [Anderson 1995], and the combination of three major factors for conducting the detection; Misuse detection, Anomaly detection and Specification-based detection [Biermann 2001], [Lippmann 1998], [Lunt 1993], [Debar 1999]. Finally, IDS can be characterized as passive or reactive depending on the type of actions taken by the system when an attack has been identified. The taxonomy of IDS [Debar 1999] along with the techniques and approaches that surround and apply to these systems are discussed by Verwoerd and Hunt [Verwoerd 2002]. A generic overview of the IDS technologies as mentioned and categorised previously are showed in [Figure 2] in an Object-Oriented approach using the Unified Markup Language (UML) notation.

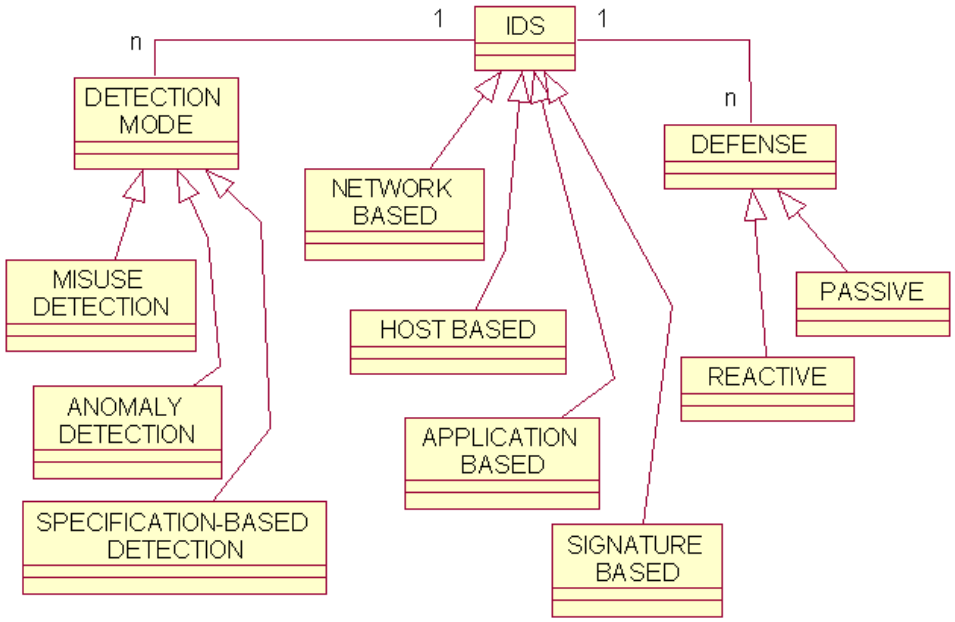


Figure 2 - Generic overview of the IDS technologies mentioned in the paper

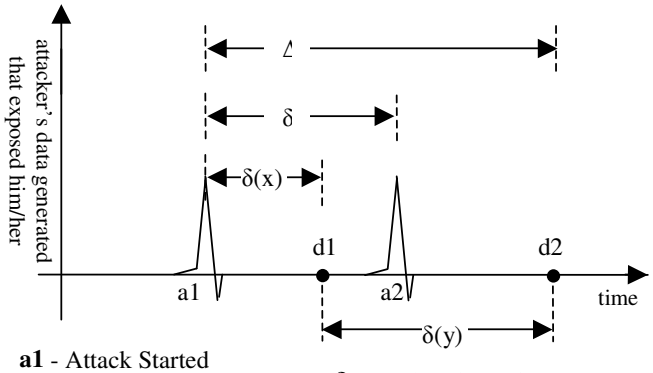
However, the solutions that are provided by the IDS and Computer Network Defense (CND) community are to be applied across a single network and/or individual systems which are usually determined by their geographical position. Some steps have been taken towards to a more distributed approach that led to systems like SnortNet [Fyodor 2000] and Prelude [Prelude 2004] but also in this case IDS are primarily developed for a single network infrastructure. Currently, ideas like the Grid for Digital Security have been introduce in order to provide a peer-to-peer based network approach [Pilgermann 2005]. Thus, this non-centralized communication architecture, which can bridge homogeneous and heterogeneous network infrastructures, can take advantage of trusted relationships in order to allow an intelligent system to perform security threat assessment.

The term "non-centralized communication architecture", mentioned previously, demonstrates the dynamic nature of the proposed system. According to the primarily design of the Security Threat Assessment [Figure 1], the Intelligent Engine resides behind the node "University E". This is not why the Threat Assessment process is taking place within the particular university but because this is the university that is being under attack at the moment. So, if we assume that for some reason a number of universities are under a Distributed Denial of Service (DDoS) attack all the valuable details about the origins of the attack will be passed directly to all the nodes of the infrastructures in order to keep them updated. Each university has an Intelligent Engine to perform threat assessment and which will take all the necessary actions. The results of the assessment can then be passed onto the other nodes instantly. Consequently, universities that are not yet under attack will be in a position to expect similar network traffic and therefore to prevent and block such an attack. However, this type of threat assessment requires it to take place in real time and not after the attack has finished or the entire infrastructure has been penetrated by malicious attackers. On the other hand, the exploitation of a novel vulnerability can be avoided due to the classified object-oriented architecture of network events [Morakis 2003a], [Morakis 2003b] and decision making by using vulnerability trees [Vidalis 2003], used by the Intelligent Engine, the defending system could have a notional understanding of the network traffic and security incidents.

Summarizing the above, combing all the techniques and technologies mentioned up to now, a large infrastructure could not only be able to be correctly administrated but also to have a notional understanding of the security incidents occurring on the network. Thus, the elevation from static IDS into more reliable dynamic IDS can be considered more efficient to stand against attack vectors like targeted (knowledgeable attacker by using social engineering, new unknown vulnerabilities like zero day exploits) and untargeted attacks (known vulnerabilities not yet patched).

The need for Real – Time Threat Assessment

The real time threat assessment has two very important goals. The first goal is to minimize the time from the moment an attack actually started until the moment our defense system is able to identify it as an attack. The second goal which we are trying to achieve is to minimize the amount of time that is essential by our system to take any required actions or deploy a set of countermeasures before the attack has finished. Explaining the attack – response timeframes we should consider the following figure [Figure 3].



a1 - Attack Started
a2 - Attack Finished
d1 - Detected Attack
d2 - Deploy Countermeasures

δ - Lasting time of an attack
Δ - Timeframe for the moment an attack detected until the moment the attack was blocked.

Figure 3 – Explaining the attack timeframes

The capital letter delta (Δ) represent the total time from the beginning of an attack until the moment we took any required actions to stop it. However if Δ> δ(y) we cannot assume that the countermeasure put in place has prevented the attacker for gaining what he/she wanted, where δ(y) is the time between the moment we realized that we are under attack until the moment that we took any necessary actions which deployed countermeasures to avoid being harmed. As motioned previously there are two goals to be achieved in order to successfully protect the infrastructure. According to the figure these are represented by the timeframe δ(x) and the second goal by the timeframe δ(y).

This can be achieved using two approaches. The first approach which is obvious is to minimize both timeframes in order to achieve the required result. The second approach is to stretch the timeframe on the

figure represented by the value δ. The value δ, as can be seen in the figure, represents the amount of time that an actual attack is taking to be completed. An attack can be considered that it has been completed when we have successfully prevented and blocked it or it has somehow achieved its goal. The second approach of expanding the time, that an actual attack will take, is quite trivial. However, it can be done by using honeypots, virtual or not, but there are still ethical issues surrounding this approach. On the other hand, minimizing the timeframes δ(x) and δ(y), where $\Delta = \delta(x) + \delta(y)$, what we are actually trying to achieve is to minimize Δ up to the point where $\Delta < \delta$. In order to achieve that we could minimize δ(y) combining today's existing technologies (e.g. Hardware Components, Processing Power, Parallel Processing, Artificial Intelligence and Software Advances). Thus, even if we try to minimize the timeframe δ(y) it won't be useful if $\delta(x) < \delta$. The timeframe δ(y) depends on continually researched sectors, so, in the future as these technologies evolve, the amount of time required for δ(y) will always be shrinking. Consequently, the only timeframe left to be compressed is δ(x). Again, by combining technologies, in order to invent an Intelligent Engine, it is possible to minimize this time. One very important key point to minimize this timeframe is to provide our defending system with "prior intelligence". This prior intelligence is provided by every little component that our large infrastructure is consisting of by the interchanging of information. Thus, even if someone succeeds in compromising successfully a machine and even if the impact of the attack is critical the rest of the important components of our system will be instantly aware of the danger and they will try to defend themselves, because now they have prior intelligence concerning this particular attack.

Because δ and Δ are totally depending from each other we have some limitation on the scale that we can minimize Δ by the nature of the problem. On the other hand δ can be as small as it can be but depending on the nature of the attack (e.g. type and target service or component) and how small δ is we could profile the attacker. Thus, not only will we be aware of the ongoing attack but also have a notional understanding of similar attacks. Therefore, no matter how clever an attacker can be we are setting our defence mechanisms to become smarter and smarter by understanding what the attackers are trying to achieve.

Performing Real – Time Threat Assessment

Incontrovertibly, such a system demands detailed research to be undertaken that will cover all the minor and difficult aspects of the project. As mentioned earlier the system will be applied across homogeneous and heterogeneous infrastructures. This is the first task that we should try to specify the easiest, most efficient and reliable way of bridging such networks and especially without having to reconfigure them. Ongoing research in this area has shown that it is possible to take audit data in various formats from various sources and unify the information gathered [Avourdiadis 2005]. Consequently this allows the information to be stored in a single database schema for further analysis and data mining.

The Intelligent Engine responsibility will be to perform real-time threat assessment which will be a combination of designing techniques and software implementations. This will consist of a classification database which categorises network events in an object-oriented form, smart load balancers, combination of the state of the art IDS technologies, distributed processing through XML documents, multi-processing architecture, parallel analysis through clustering and probably artificial intelligent methodologies [Fragkos 2005]. The primary and ultimate goal to be achieved through the real-time threat assessment is to minimize as much as possible (or up to the point of extinction) the false negative and false positive alerts. Consequently, the process of identifying an attack (even in its primarily state), analyze it and finally deploy a set of countermeasures, will become slightly easier.

The ongoing research conserving the real-time threat assessment is currently at the stage of expressing it as near real-time [Fragkos 2005]. This is because the stage where the events are being analyzed requires further research in order for the system to be able to have a notional understanding of the network traffic that it is monitoring and of course it must be capable of predicting the attacker's next step. Finally, after passing the point where the system has developed an "idea" of what and why it is being attacked, it should be capable of reporting it (at least the impression the system has about the active/on-going attack) in a legitimate way. The reporting mechanism will be the fundamental stone for the proper countermeasure engine to be developed and applied. The process of analysis and deployment of countermeasures, as we have just described, represents the timeframe δ(y) [Figure 3]. This timeframe must be kept as narrow as possible as it can be. The logical step to take in order to achieve this is by minimising the amount of time spent on processing the information. The only way to achieve this is with the correct use of optimized algorithms processed on high processing powered machines (i.e. multi processor computer, Beowulf clusters).

Conclusion

The purpose of the paper is to introduce a wider idea concerning today's intrusion detection and defence mechanisms, used by systems with critical importance. Despite any advantages currently available, IDS lack in performing threat assessment due to the incapability of merging, comparing, and analyzing network events form disparate heterogeneous sources in real time. Moreover, they do not offer a reliable way of deploying them across large infrastructures.

Securing the existing critical-importance infrastructure should be treated as a primary consideration. The impact of the loss of such a system could be major, catastrophic and even cause the loss of human life. Trying to provide answers to the questions set forth in the introduction it should be mentioned that the realistic picture is that network infrastructures will always be under attack, no matter what. The reason why we cannot yet deal efficiently with these attacks resides with our incapability to merge and analyze the appropriate generated information in order to defend the infrastructure in a correct manner. The problems exist in performing security threat assessment should not be considered as technological problems. But, as we have seen throughout the paper it has to do with the appropriate combination and proper use of specific components. Consequently, we will manage to develop the architecture under question and implement intelligent mechanisms that will take advantage of currently available resources (software, hardware etc). The final goal is to reach up to the point where real-time security threat assessment will become a reality and thus, a computer infrastructure will be capable of automating the process to think and protect it self.

References

Anderson D., Lunt T., Javitz H., Tamaru A., Valdes A., (1995) Next-generation Intrusion Detection Expert System (NIDES): A Summary, SRI-CSL-95-07, SRI International

Avourdiadis, N., Blyth, A., (2005), Data Unification and Data Fusion of Intrusion Detection Logs in a Network Centric Environment, 4th European Conference on Information Warfare and Security, University of Glamorgan

Biermann, E., Cloete, E. and Venter, L. (2001). A Comparison of Intrusion Detection Systems. Computers & Security 20 (8): 676-683.[12]

Blyth A. (2003) An XML-Based Architecture to Perform Data Integration and Data Unification in Vulnerability Assessments, Information Security, 8 (4) 14-25

CERT® Coordination Center, CERT Coordination Center Statistics 1988-2003, Available at: http://www.cert.org/stats/cert_stats.html

Debar H., Dacier M., Wespi A., (1999) Towards a taxonomy of intrusion detection systems, Computer Networks 31 (8): 805-822

Fragkos G., Blyth A., (2005), Architecture for Near Real-Time Threat Assessment using IDS Data, 4th European Conference on Information Warfare and Security, University of Glamorgan

Fyodor Y. (2000) Snortnet - A distributed Intrusion Detection System, Available at: http://citeseer.ist.psu.edu/fyodor00snortnet.html

Goodwin, B., (2002), Record Wave of Hacking Targets UK Business, Computer Weekly, 31 October 2002: 6

Humphry Crum Ewing (1998), Smart Procurement, The Standish Group, Available at: http://www.publications.parliament.uk/pa/cm200304/cmselect/cmdfence/572/572we12.htm

Lippmann R., Fried, D., Graf, I., Haines, J., Kendall, K., McClung, D., Weber, D., Webster, S., E., Wyschogrod, D., Cunningham, R., K., Zissman, M., A. (1998) Evaluating Intrusion Detection Systems, The 1998 DARPA Off-line Intrusion Detection Evaluation. First

Lunt, T. (1993) A survey of intrusion detection techniques, Computers and Security. 12 (4): 405-418

MoD (2001), Smart Procurement, Ministry of Defence, Available at: http://www.mod.uk/issues/sdr/procurement.htm

Morakis, E., Vidalis, A., Blyth, A. J.C. (2003a). Measuring Vulnerabilities and their Exploitation Cycle, Elsevier Information Security Technical Report, Vol. 8, No. 4

Morakis, E., Vidalis, S., Blyth, A.J.C. (2003b). A Framework for Representing and Analysing Cyber Attacks Using Object Oriented Hierarchy Trees, 2nd European Conference in Information Warfare, UK, pp235-246

NISCC (2005), National Infrastructure Security Co-ordination Centre, Available at: http://www.niscc.gov.uk/niscc/index-en.html

Prelude (2004) An Open Source Hybrid Intrusion Detection System, Available at: http://www.prelude-ids.org/article.php3?id_article=66

Verwoerd, T. and Hunt, R. (2002), Intrusion Detection Techniques and Approaches. Computer Communications, Elsevier, U.K., Vol 25, No 15, September 2002, pp1356-1365

Vidalis, S., Jones, A., (2003), Using Vulnerability Trees for Decision Making in Threat Assessment, Technical Report, University of Glamorgan.

Exploring the Role of Deception in Computer Network Defence and Information Operations Environments

Zafar Kazmi, Theodore Tryfonas, Stilianos Vidalis

School of Computing, University of Glamorgan, UK

zkazmi@glam.ac.uk tryfona@glam.ac.uk svidalis@glam.ac.uk

Abstract
The basic concept of deception has been practised perhaps since the natural world has existed but its application to an information security environment could only be witnessed since the early 1990's. Since then, different deception techniques have also been introduced to the Computer Network Operations (CNO) and Information Operations (IO) environments, in order to enhance the outcome of a specific campaign. In the recent years, a number of researchers have investigated in different deception techniques used in information security and computer based networks but more consideration is needed in exploring the strategic deception in a Computer Network Defence (CND) environment. This would help in improving the security of an organisation's Critical Information Infrastructure (CII) by enhancing the organisation's ability to survive a Computer Network Attack (CNA). This paper will outline the importance of the role that deception can play both in an IO and a CNO environment.

Keywords: Deception, Information Security, Computer Network Defence, Computer Network Operations, Information Operations.

1. Introduction

Over the last decade, the increased use of computer systems and the swift boost of the Internet were accompanied by the equal growth of computer security threats and incidents. Both, the technology and the threats related to these technologies are becoming more and more complex and therefore use of targeted deception can be advantageous in a computer systems security environment. Therefore, information and computer systems face a wide variety of threats which can result in significant damage to an organisation's vital infrastructure.

The range of threats varies from threats to data integrity resulting from unintentional errors and omissions, to threats to system confidentiality, integrity and availability from malicious intruders attempting to compromise a system. Awareness of the threats and vulnerabilities of a particular system allows the selection of the most effective security measures for that system. This includes building a strong network defence by employing physical, procedural and personnel security measures as well as deploying electronics security measures such as Firewalls, Anti-Viruses, Intrusion Detection Systems (IDSs), Access Control Lists (ACLs), deployment of deceptive techniques, etc.

Network defence is also put in place to deal with other types of attacks such as service interruption, interception of sensitive email or data transmitted and use of computer's resources. In specific, network defence is about taking measures that should reduce the likelihood of intruders breaking into an organisation's critical computer network and causing damage by reading or stealing confidential data or even modifying it in order to sabotage that organisation. Here, an organisation can be an independent establishment, or a group of government officials, or the actual government of a country, that ensures the security and stability of the critical information infrastructure of their organisation or country.

2. A Review of Concepts and Terminologies

2.1. Deception

The basic concept of deception is an ancient one, existing in nature, but the application of different deception techniques in a computer network security and information security environment emerged in the early 1990's. In specific, deception is an act of deceiving or misleading and can also be defined as "*the problematic distinction between appearance and reality*" (Rue 1994).

The deception used in military operations is defined in the United States Joint Doctrine for Military Deception as:

"*Actions executed to deliberately mislead adversary military decision makers as to friendly military capabilities, intentions, and operations, thereby causing the adversary to take specific actions that will contribute to the accomplishment of the friendly mission*" (JCS, 1996).

Deception can be considered as the creation and invocation of both offensive and defensive environments and can be employed for attacking an adversary's perception of what is actually occurring. Furthermore, deception can be applied to enhance an operation, exaggerate, minimise, or distort the enemy/opponent's perception of capabilities and intentions, to mask deficiencies, and to otherwise cause a desired outcome where conventional military activities and security measures were unable to achieve the desired result (Cohen & Lambert, 2001).

A famous Chinese General known as Sun Tzu outlined in a collection of essays called "The Art of War" that: *"All warfare is based on deception. Hence, when able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near. Hold out baits to entice the enemy. Feign disorder, and crush him. If he is secure at all points, be prepared for him. If he is in superior strength, evade him. If your opponent is of choleric temper, seek to irritate him. Pretend to be weak, that he may grow arrogant. If he is taking his ease, give him no rest. If his forces are united, separate them. Attack him where he is unprepared, appear where you are not expected."* (Rongstad, 1996), (Sun, 1983), (Griffith's, Undated).

Moving on, the deployment of effective deception can also be an important element of information and computer based system's security. In the past, different deception techniques have been introduced to play their role in information security and to secure a computer based network. For instance, deployment of Honeypots and Honeynets as shown in figure 1 below (Spitzner 2003), in a computer based network can lead to the discovery of an attacker's movements and allow the network to be secured against the attacker's next offensive move and strategies.

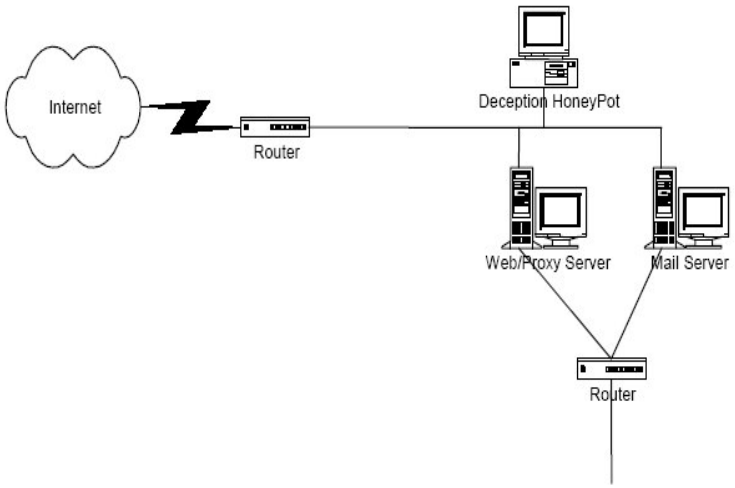


Figure 1: Typical Honeypot Deployment (Spitzner, 2003)

In specific, Honeypots are systems designed to be appeared as fully functioning elements of the infrastructure, placed at an appropriate location on the network where all inbound and outbound traffic is captured and monitored, providing a secure and controlled environment to allow attackers to access them (Gupta, 2003), (Spitzner, 2003).

Similarly, the art of deception can also be deployed in a Computer Network Operations (CNO) environment and a number of different deception techniques have also been introduced to different IO campaigns, resulting in enhancement of the outcome of the actual operation. Therefore, it is likely that in the near future,

deception implemented through high-tech means will play an increasing role in both IO and CNO environments.

2.2. Critical Information Infrastructure (CII)
All critical infrastructures including transportation, finance, water, electric power, public telephone network, the Internet, and terrestrial and satellite wireless networks for a variety of information management, communications, and control functions are increasingly dependent on the evolving Information Infrastructure of a country.

Similarly, an organisation has its own CII including financial controls, information systems, computer network systems, etc. and therefore, security of an organisation's information infrastructure (II) is vital. Here, it is essential to mention that the CII can be seen as the subset of the II as shown in figure 2. It would therefore be a good idea to investigate information infrastructure before actually exploring the CII.

In specific, *II is an integrated system comprising of computing, communications, and the actual information stored within the system as well as the people who use and operate this technology* (Busuttil & Warren, 2003).

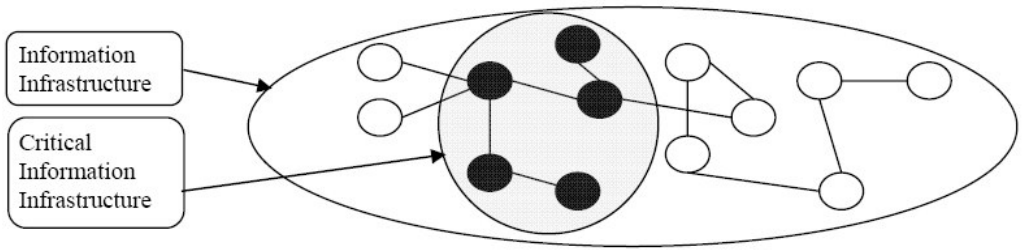


Figure 2: The relationship between II & CII (Busuttil & Warren, 2003)

CII on the other hand, consists of *the minimum amount of human and technological entities within the information infrastructure which needs to be in fully functioning state for an organisation to have information based supports for its business activities* (Busuttil & Warren, 2003). Here it is important to mention that the protection of both II and CII includes securing and defending the basic facilities services, information systems itself and more importantly securing the actual elements needed to ensure successful operation of an organisation's information systems.

2.3. Computer Network Operations (CNO)
Computer Network Operations (CNO) can be defined as a combination of Computer Network Attack (CNA), Computer Network Defence (CND) and Computer Network Exploitation (CNE). It would therefore be adequate to gain initial understanding of CNA, CND & CNE, in order to appreciate the concept of CNO:

Computer Network Attack (CNA) can be described as the *"Operations carried out using computer hardware or software, or conducted through computers or computer networks, with the intended objective or likely effect of disrupting, denying, degrading, or destroying information resident in computers and computer networks, or the computers and networks themselves"* (United States Joint Forces Command Glossary, Undated).

The Computer Network Defence (CND) on the other hand, is "*the measures taken to protect and defend information, computers, and networks from intrusion, exploitation, disruption, denial, degradation, or destruction*" (United States Joint Forces Command Glossary, Undated).

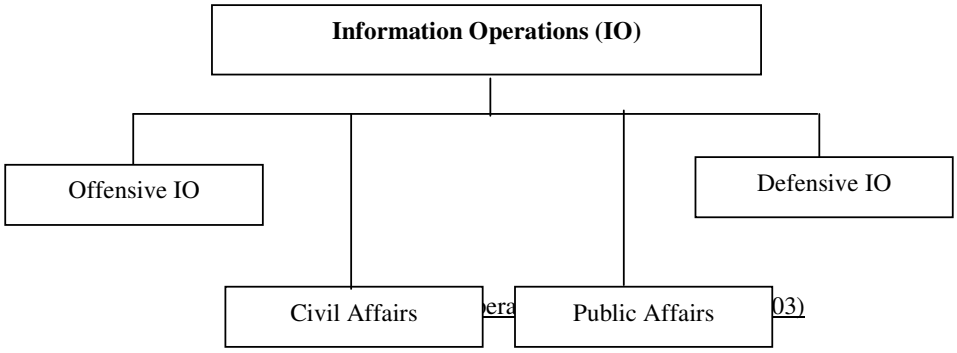
Finally, the Computer network exploitation (CNE) can be defined as "*the intelligence collection and enabling operations to gather data from target adversary automated information systems (AIS) or networks*" (United States Joint Forces Command Glossary, Undated).

2.4. Information Operations (IO)
To date, there are over seventeen different definitions of Information Operations (IO). The considered definition of IO for the purpose of this research is the one stated by the Qinetic, a major United Kingdom's

defence contractor. According to Qinetic, IO is *"the strategic planning and coordination of activities necessary to protect an organisation's information"* (QinetiQ, 2003).

Here it is essential to mention that defensive IO, unlike offensive IO, are carried out in order to protect and defend information system by introducing, integrating and co-ordinating policies, procedures, personnel and technology (Jones & Ashenden, Undated).

Figure 3 below outlines different IO categories:



Deception can also play an important role in a successful IO campaign. According to the United States Joint Chiefs of Staff (JCS) Memorandum of Policy (MOP) 116, *"Military deception has proven to be of considerable value in the attainment of national security objectives, and a fundamental consideration in the development and implementation of military strategy and tactics" (JCS Memorandum of Policy (MOP) 116 [10]) (Unnamed, Undated).*

The MOP then further states that "*The development of a deception organization and the exploitation of deception opportunities are considered to be vital to national security" (JCS Memorandum of Policy (MOP) 116 [10]) (Unnamed, Undated).*

Furthermore, the relationship between deception and IO can be defined as the one based on psychological component (Douglas, 1998). It is therefore considered that Psychological Operations (PSYOPS) can be an integral part of a successful military operation. This is evident from the United States Department of Defence's (DoD) statement *"PSYOPS are a vital part of the broad range of United States diplomatic, informational, military, and economic activities"* (JCS, 2003).

Finally, the IO battle space includes and not limited to:

1. Corporate Level: Netspionage, Sabotage, Destruction of magnetic media, Computer theft, Competitor trash capture & analysis.

2. Personal Level: E-commerce fraud, Spoofing, E-mail harassment, Spamming, Card theft.

3. The Role of Deception

Deception is an ancient art, and an art it is indeed, as noted in many sources, one of them being Dearth (Campen and Dearth, 1998). It goes back to the 10th century BC when King Solomon said: *"A wise man has great power, and a man of knowledge increases strength; for waging war you need guidance, and for victory many advisers."* The more information one has the better he will be able to assess a situation in taking advantage of certain variables for achieving information superiority.

The following table was taken from Waltz (Waltz, 1998).

IW Model Layer		Function	NETWAR
Offence	Perceptual	Manage perception, Disrupt decision processes	PSYOPS, Deception
	Information	Dominate information infrastructure	NETOPS
	Physical	Break things…, Incapacitate/kill people	Physical destruction
Defence	Perceptual	Protect perceptions and decision-making processes	Intelligence, Counterintelligence
	Information	Protect information infrastructure	INFOSEC
	Physical	Protect operations, protect people	OPSEC

Table 1 - Taxonomy of information operations (Waltz, 1998, p.208)

NETWAR has been defined by (Waltz, 1998) as:
"The information-related conflict waged against nation states or societies at the highest level, with the objective of disrupting, damaging, or modifying what the target population knows about itself or the world around it."

Through deception we can manage our adversary's perception and disrupt his decision-making processes. These processes feed into his (the adversary's) defensive INFOSEC processes which when disrupted will allow the success of our offensive NETOPS (Waltz, 1998) that will ensure out information superiority. In specific, *"Information superiority is the capability to collect, process, and disseminate an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same."* (Waltz, 1998)

Knowledge on the other hand, is intelligence and if we are able to disrupt the intelligence and counterintelligence operations of our adversary then we can achieve information superiority. The common operation is deception and agreeing with Cohen (Cohen, 1998) it is believed to be the future of IO and Information Security.

Deception allows subduing the enemy without fighting. Fighting cost resources and resources cost money, which is a very scarce resource… Security though should not be an expensive commodity. The weakest link of a game should be able to afford the same level of security the strongest link have. It is accepted (Vidalis, 2004) that the weakest link does not get thrown out of the game, it destroys the game altogether. Furthermore, fighting is usually a reactive action. It has been proven though (Vidalis, 2004) that in the area of security reacting is expensive, instead we want to be proactive and somehow prevent attacks from happening or minimising/nullifying the impact of the threats.

The purpose of deception is to surprise the adversary. If the threat agent is in a state of surprise the outcome can be twofold: either the defenders have time to react and deploy the necessary countermeasures (or finely tune the existing ones), or the threat agent will call off the attack and return to the information gathering process in order to re-examine his plan of action.

Agreeing with Mitnick (Mitnick and Simon, 2002), technology has allowed for an increased capability for information gathering, but perceptions and the nature of decision-making have one common vulnerability: the human factor. Humans sit behind monitors, typing and/or communicating commands. Humans are in charge of automated procedures and can shut them down if they perceive that something is wrong and that the computer reactions do not make sense. Of course there are examples of computers being in charge, one being the flooding system in the port of Amsterdam, but the author is yet to meet a general that doesn't like to have control of everything, or more of the point, that hasn't got control of everything. Under the same perspective the above applies to network administrators. The author has yet to meet a professional network administrator that does not like to have complete control over his system (which constitutes a major vulnerability).

The responsibilities of the network administrator is summarised in the following list:

- Design a network which is logical & efficient
- Deploy large numbers of machines which can be easily upgraded later
- Decide what services are needed
- Plan and implement adequate security
- Provide a comfortable environment for users and keep them happy…
- Develop ways of fixing errors and problems which occur
- Keep track of, and understand how to use, technology

By designing a logical network though, the administrator makes the life of the threat agents easier, as they can follow the same logic and enumerate the infrastructure. Deception can be used to hide the real computers amongst the false. By having easily upgradeable computers the administrator possibly introduces a critical threat against his infrastructure. Should the upgrade procedures get compromised then threat agents will be able to launch catastrophic active attacks. Again deception can be used to masquerade the procedures and/or produce confusion about what is real. Some would argue that you can never have enough security, a statement that has been argued from the threat and risk assessment professionals. Agreeing with Mitnick (Mitnick and Simon, 2002) though the network administrator will always have to fear the users of his system. System users are probably the bigger vulnerability of that system as they are susceptible to social engineering attacks.

If we consider that deceiving our own users is acceptable then deception can offer a solution to the social engineering vulnerability. To summarise, deception can be used in two ways for ensuring security in a computing infrastructure:

- Simulating: showing the false, drawing attention away from the real
- Dissimulating: hiding the real, producing confusion about what is real

Moving on to the basic definition of IO as described by the Qinetic: *"the strategic planning and coordination of activities necessary to protect an organisation's information"* (QinetiQ, 2003). Here, an "organisation" can be referred to a number of different types of establishments such as a country, a country's government officials or military forces or even an enterprise whether in the private sector or owned by a country's government. For the purpose of this research, the word "organisation" would refer to establishments both private and government organisations such as Telecommunications companies, Electric & Water supply companies, Emergencies services, Air Traffic Control services etc.

Here, it is important to mention that deployment of Psychological Operations (PSYOPS) could make IO more efficient and may also help in achieving the desired goal more rapidly. "*PSYOPS support IO by developing products that develop understanding and favourable attitudes of the local populace toward the peace operation force; gain local support for the military effort; and, help attain the objectives of the friendly force"* (IWS, Undated). In specific, PSYOPS are *"Planned operations to convey selected information and indicators to foreign audiences to influence their emotions, motives, objective reasoning, and ultimately the behaviour of foreign governments, organizations, groups, and individuals. The purpose of psychological operations is to induce or reinforce foreign attitudes and behaviour favourable to the originator's objectives"* (IWS, 2004).

Another simple definition of PSYOPS is "*the use of communications (such as propaganda) and actions intended to mislead to influence the perceptions, motives and emotions of the enemy*" (Yoshihara, 2001).

The deployment of PSYOPS can be witnessed from the latest war against Iraq when United States and its Coalition forces dropped warning leaflets and radios over southern Iraq. Along with many leaflets, one of the leaflets pictured an armed Iraqi soldier at the far right and an anti-aircraft gun at the left (Friedman, 2003). *The anti-aircraft gun is firing at the allied aircraft but the shells are actually exploding far behind and therefore not hitting the actual aircraft. The leaflet also has text written in Arabic and says, "Before you engage Coalition aircraft, think about the consequences". The back of the actual leaflet showed an Iraqi soldier surrounded by smoke and under the attack of Coalition forces. The leaflet also showed an Iraqi woman with two infants and the text on the back reads, "Think about your family and do what you must to survive*" (Friedman, 2003).

These types of leaflets were successful to a certain extent (Friedman, 2003). It is considered that this was due to the believable and convincing emotional messages which played their role in disarming Iraqi forces and therefore proved the benefit of PSYOPS in a successful military manoeuvre. The limited success could be because Iraqis are nationalistic and perhaps found it hard to accept that outsiders (i.e. Coalition forces) would really be interested in their national well being. Another major reason for the limited success of these PSYOPS could be that Israel is one of the United States closest allies and Iraqis, being Muslims, consider that as a threat to their country as well as to their religion, which is an integral part of their lives.

Moving on to a threat assessment carried out by the United States Navy, some nation states such as China, Russia and India are reported to have developed different policies of preparing for a cyber-warfare and are

engaged in rapidly developing their IO capabilities (Hildreth, 2001). The report further indicates that countries such as Iran, Syria, Libya and North Korea have some IO capabilities whereas other countries such as France, Japan and Germany are comparatively advanced capabilities (Hildreth, 2001). Here, it can not be assumed that all these nations are only investigating defensive IO capabilities and not exploring the offensive IO capabilities. It is considered that in order to operate effectively, all nation states will need to be equipped with defensive IO capabilities in order to protect their important infrastructure.

It is also considered that misleading information and deceptive tactics can play an important role in a successful campaign ensuring a desired completion of a specific military operation and enabling you to monitor your opponent's moves. This can be seen from the United States DoD's statement about the importance of PSYOPS in a military operation, that *"PSYOPS are a vital part of the broad range of United States diplomatic, informational, military, and economic activities"* (JCS, 2003)

Another excellent example of application of deception techniques in an IO campaign is when United States military started its email campaign of urging Iraqi military and civilian leaders to take over Saddam Hussein's regime (Friedman, 2003). One of the several emails stated that "*Iraqi chemical, biological and nuclear weapons violate Iraq's commitment to agreements and United Nations resolutions. Iraq has been isolated because of this behaviour. The United States and its allies want the Iraqi people to be liberated from Saddam's injustice and for Iraq to become a respected member of the international community. Iraq's future depends on you*" (Friedman, 2003).

Iraqi authorities responded to that threat by blocking the emails in order to ensure that the messages do not spread throughout the country (Friedman, 2003). This example shows deployment of deceptive techniques by the United States military in an attempt to win the hearts and minds of Iraqi military and civilian leaders. Because, one could argue that there was no strong evidence of Iraq having chemical, biological and nuclear weapons and perhaps was based on assumptions by the United States intelligence. Hence, using this issue in order to gain objectives could be considered as misleading and therefore justifies with the definition of deception.

It is considered that there was no strong evidence that United States military was intended to turn Iraq into a well developed country and perhaps there were other incentives. This also indicates to the author that the US military deployed deceptive techniques in order to enhance their military operations in Iraq. Here, it is also considered that if those PSYOPS and deceptive tactics had not been deployed in Iraq by the US military, there may have been more resistant from the Iraqi forces since it was evident during the war that a number of Saddam's forces surrendered.

Iraqi forces also employed deceptive tactics before and during the war against Coalition forces in order enhance their operations. Saddam Hussein stated on the Iraqi state television that majority of the leaflets dropped by the Coalition forces were burned by the Iraqi people to show that they did not trust the Coalition forces (Friedman, 2003). This may have been done in an attempt to make the Coalition forces believe that the leaflet campaign was not effective at all or even to convince more and more Iraqi people to follow the same trend as others (i.e. not to believe the context of the leaflets and set them on fire.

Finally, an article was published in Washington Monthly by Joshua Micah Marshall and stated that the United States would deal with Syria and Iran after it had finished dealing with Iraq (Marshall, 2003). The article further mentions how the whole issue of war with Iraq was full of deceptive information provided by the United States (Marshall, 2003). This clearly shows that in this hi-tech age of information systems, deception can play an increasingly important role to achieve desired objectives when deployed in an IO campaign.

5. Discussion

The use of deception in military operations is as ancient as the existence and understanding of actual war. As long ago as 1469 BC**,** *during the reign of Thutmose III, the Egyptians used different deception techniques to fool their enemies, and pass into Syria through an unsecured route* (Sun, 2002).

Similarly, employment of different deception strategies in an IO environment have been part of a successful military operation for a long time. The range of IO tools include: malicious software, denial of service, spoofing, cryptology, electromagnetic pulse weapons, destructive microbes and psychological operations. When a combination of these tools is applied, together with a suitable deception method, a successful Information Operation can be achieved. There are numerous examples of the use of different deception techniques in an IO campaign in order to enhance the operation such as the ancient Homer's tale of the

Trojan-Horse demonstrating an important role that deception played in warfare at the dawn of European history (Sun, 2002).

Even World War II provides a number of examples of the deployment of deception techniques (Sun, 2002). For instance, when British deception misled the German intelligence by making them receive wrong results of their targeting of V-1 and V-2 missile attacks (Unnamed, Undated, www.2worldwar2.com). Basically, *"The British intelligence used captured German agents to transmit to Germany, the lists of correct locations where the German missiles fell, but with mixed dates. The unsuspecting Germans compared those lists to their own log of missile attacks targeting data, and used the differences between the lists for aiming corrections. This misleading information made Germans to increase their aiming error instead of decreasing it, which resulted in Britain saving many innocent lives"* (Unnamed, Undated, www.2worldwar2.com). Here, it is obvious that the use of deception enabled Britain achieving its military objectives and if the deception techniques had not been used, the outcome of this operation may have been different.

Based on the research carried out for this paper, it can be stated that a number of different deceptive techniques can be introduced to an organisation's CNO, which can play their important role in defending against, as well attacking an adversary. Here, the main issue is to ensure that effective deception is deployed in order to achieve a desired outcome. Therefore, a method of strategic deployment of deception in a CND environment may be introduced in order to increase an organisation's information infrastructure's security and reliability. Here, it is important to mention that CND is a vital part of a CNO campaign, as it is also evident from the definitions of CNO and CND, outlined in the preceding literature.

To date, there are no methodologies available which would lead to the strategic deployment of deception in an IO or a CNO environment. Similarly, there are no methodologies or models available that would allow targeted deception in a CND environment which may enhance the ability of an organisation's CII, in order to survive a CNA. There is a generic framework available for deception, designed by Fred Cohen (Cohen, Undated), but it does not offer effective and suitable deployment of different deceptive techniques for a specified CNO operation. Similarly, a cognitive model for exposition of human deception and counter-deception was also introduced by D. Lambert back in 1987 (Lambert, 1987). The model is based on developing a basic understanding of human deception which would then lead to a comprehensive development of framework for organising deception principles and examples. Just like Fred Cohen's frame work for deception, D. Lambert's model for deception is generic and does not allow targeted deception in a CNO environment. Therefore it would be sufficient to state that the development of a methodology that would allow strategic deployment of deception in a CNO environment would be an immense achievement.

The planning process for a targeted deception operation has to be a backwards-planning process. This means that the desired end-result would become a starting point and would then derive the actual target and achievement. This is where the idea of deriving a methodology for deploying deception in a CND environment originates from and further work will be carried out to achieve this goal. Figure 4 below outlines the planning process for deception.

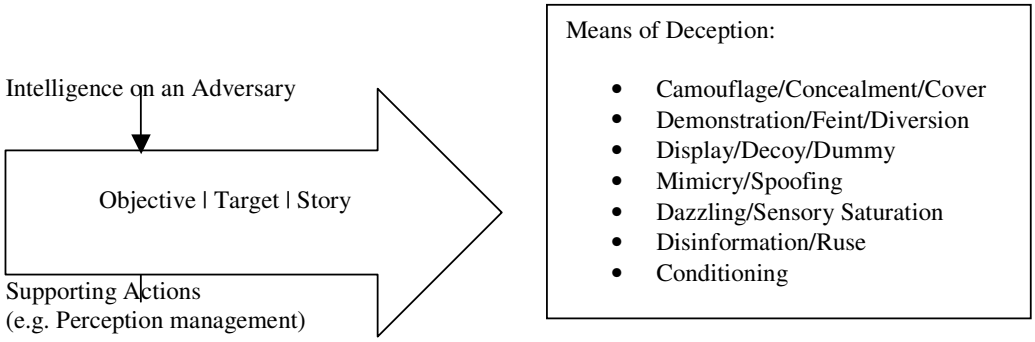


Figure 4: Deception Planning Process (Gerwehr and Glenn, 2003, p.26)

Although, the methodology will be aimed to be developed for a CND environment, it is obvious that it may lead to deployment of some offensive deception techniques since defensive deception techniques can sometimes be seen as offensive procedures, depending on the actual situation. Basically, the starting point for planning of developing the methodology would be to outline the targets that an organisation may aim to achieve. There are a number of different targets that an organisation may intend to achieve including the defence of its CII against an intruder looking to compromise the organisation's CND.

Similarly, the information gathering process would also play an important role in development of the methodology in discussion. Furthermore, it can be predicted that the initial part of the developed model would be based on collecting information about the intruders / attackers. Although, this research is in its very early stages of planning of development but the critical review of the literature search has led to establish some understanding and basis for the methodology. The following deception techniques could be useful when it comes to gathering information of an attacker with an intention of attacking an organisation's CII by defeating its CND:

- Concealment or hiding
- Camouflage (hiding movements from the intruder by artificial means)
- False and planted information (Misinforming "letting the intruder have the information that may hurt the intruder and may lead to learn attacker's next move")
- Displays ("techniques to make the enemy see what is not actually there")
- Ruses ("tricks, such as displays that use enemy equipment and procedures")
- Insight ("deceiving the attacker by out thinking him")

If a combination of these techniques is deployed in a systematic manner, an attacker's movements may be directed through a series of different deception techniques that induces the attacker into deceived states. Furthermore, the initial part of the developed method may also include ways of assessing the skills of a likely intruder. This can be achieved by considering the nature of the business of an organisation and then try to predict the type of attacker that organisation is likely to attract. Obviously, the research would not restrict to just these objective since other factors will be involved which will be identified as the research progresses in the future.

6. Conclusions

Deception can be considered as a vital element of both information security and computer based systems security and therefore can play an increasingly important role to achieve desired objectives when deployed in an IO or a CNO environment. It is also concluded that targeted deception should:

- reinforce enemy expectations
- have realistic timing and duration
- be integrated with operations
- be coordinated with concealment of true intentions
- be imaginative and creative

Also, deception is an essential component of military tactics and is becoming an integral part of a successful IO campaign. The significance of PYSOPS in IO, as witnessed in the first Iraq war, reflects the importance of deploying the appropriate deceptive techniques in order to enhance the operation.

Although, a number of information and computer systems security related frameworks are available out there but the organisations do not have enough guidance in the field of CND and CII protection. Furthermore, there are no methodologies available that would enable an organisation to employ strategic deception in order to increase the security of its CND. This could be due to the fact that deployment of deception in a CND is still in its infancy and a lot of research can still be carried out to achieve a milestone in this area of CNO.

Finally, the network defence is becoming increasingly important in the field of this high-tech and fast moving information technology. Therefore, one of the goals in network security should be to improve defences through the use of deception proactively against a target such as an intruder aiming to target the network defence of an organisation's CII by compromising its CND. Hence, it would be beneficial to design a methodology that would lead to the deployment of strategic deception in CND environment.

7. References

Ashenden, D. & Jones, A. (Undated). Re-Interpreting Information Operations for the Private Sector, 2nd European Conference of Information Warfare.

Busuttil, T. & Warren, M. (2003). A Review of critical Information Infrastructure Protection within IT Security Guidelines, 4th Australian Information warfare and IT security Conference 2003.

Campen, A. D. and D. H. Dearth (1998). Cyberwar 2.0: Myths, Mysteries and Reality. Fairfax, Virginia, AFCEA International Press

Cohen, F & Lambert, D. (2001). A Framework for Deception, [online], http://all.net/journal/deception/Framework/Framework.html

Cohen, F. (1998). "A Note on the Role of Deception in Information Protection." Computers & Security 18 November 1998 17(6): 483-506

Friedman, A. H. (2003). No-Fly Zone Warning Leaflets to Iraq, [online], http://www.psywarrior.com/IraqNoFlyZone.html

Gerwehr, S. & Glenn, R. W. (2003). Unweaving the Web: Deception and Adoption in Future Urban Operations, Rand, Santa Monica.

Griffith's (Undated). Sun Tzu & the Art of war Applied to Portfolio & Risk Management: From Griffith's

Translation: Deception and Shaping, [online], http://www.strategies-tactics.com/suntzuchp1.htm

Gupta, N. (2003). Improving the Effectiveness of Deceptive Honeynets through an Empirical Learning Approach, [online], http://www.infosecwriters.com/text_resources/pdf/Gupta_Honeynets.pdf

Hildreth, A. S. (2001). CRS Report for Congress: Cyberwarfare, [online], http://www.fas.org/irp/crs/RL30735.pdf

IWS. (2004). Definition of Psychological Operations, [online], http://www.iwar.org.uk/psyops/

IWS. (Undated). Chapter Three “Operations”: Psychological Operations, [online], http://www.iwar.org.uk/iwar/resources/call/iochap3.htm

JCS. (1996). Joint Doctrine for Military Deception, Joint Pub 3-58, [online], http://www.fas.org/irp/doddir/dod/jp3_58.pdf

JCS. (2003), Doctrine for Joint Psychological Operations: Overview, Joint Publication 3-53, [online], http://www.iwar.org.uk/psyops/resources/doctrine/psyop-jp-3-53.pdf

Lambert, D. (1987). A Cognitive Model for Exposition of Human Deception and Counterdeception), [online], http://jps.lanl.gov/vol1_iss1/5Cognitive_Model_of_Deception.pdf

Marshall, M. J. (2003). Practice to Deceive, [online], http://www.washingtonmonthly.com/features/2003/0304.marshall.html
Mitnick, K. D. and W. L. Simon (2002). The Art of Deception. Indianapolis, USA, Wiley Publishing

QinetiQ. (2003). Information Operations (IO), [online], http://www.qinetiq.com/home_enterprise_security/datasheet_index.Par.0002.File.pdf

Rongstad, R. (1996). Sun Tzu: The Art of War, [online], http://vikingphoenix.com/public/SunTzu/suntzu.htm

Rue, L. (1994). By The Grace ff Guile: The Role of Deception in Natural History and Human Affairs. New York: Oxford University Press.

Spitzner, L. (2003). Honeypots - Tracking Hackers. Boston: Pearson Education Inc.

Sun, T, (1983). The Art of War, Translated by James Clavell, Dell Publishing, New York, NY.

Sun, T. (2002). Library Notes: Deception, United States Naval War college, Vol. 31, No. 3, [online], http://www.nwc.navy.mil/library/3Publications/NWCLibraryPublications/LibNotes/libdeception.htm

(United States Joint Forces Command Glossary, Undated). [online], http://www.jfcom.mil/about/glossary.htm

Unnamed. (Undated). The Mechanisms of Defeat Examples from World War 2, [online], http://www.2worldwar2.com/defeat-examples.htm

Vidalis, S. (2004). Threat Assessment of Micro-payment Systems, School of Computing. Pontypridd, University of Glamorgan

Waltz, E. (1998). Information Warfare. Norwood, USA, Artech House

Yoshihara, T. (2001). Chinese Information Warfare: A Phantom Menace or Emerging Threat, [online], http://www.iwar.org.uk/iwar/resources/china/iw/chininfo.pdf

Denial of safety critical services of a Public Mobile Network for a critical transport infrastructure

Ester Ciancamerla, Michele Minichino
ENEA CR Casaccia, sp Anguillarese 301, 00100 Roma, Italy
{ciancamerlae, minichino}@casaccia.enea.it

Abstract. Denial of service measures of the Public Mobile Network (PMN) of a Tele Control System prototype, developed inside the EU SAFETUNNEL Project, are predicted by using stochastic mo dels. Modern society increasingly depends on communication networks, even public and mobile, usually born for not critical aims and nowdays more and more used for safety critical aims. In automotive domain, Tele Control Systems (TCS), based on communication networks, are proposed in prototypal fashion, for protection of critical transport infrastructures, such as long monotube alpine road tunnels. TCS typically consists of one Tunnel Control Centre (TCC), allocated nearby the infrastructure, interconnected to vehicles to be controlled, by a wireless communication network, even public (PMN - Public Mobile Network). PMN is the heart of TCS and makes it more vulnerable due to many factors, such as complexity, mobility of nodes, response time and public access to the network. The paper focuses on a PMN that supports both Global System Mobile (GSM) and General Packet Radio Service (GPRS) connections, for voice and data transmissions between instrumented vehicles and TCC. Particularly, we present PMN models, based on Stochastic Activity Networks formalism, to compute denial of service measures of voice and data connections, in the framework of validation by modelling of the related TCS. Denial of service models are composed models, each one joining two modular sub models, respectively representing PMN availability and performance aspects.

Keywords denial of service, dependability, stochastic models, public mobile networks, global system mobile, general packet radio service

Introduction

A Tele Control System is under prototypal development in the frame of the SAFETUNNEL EU Project [1]. The Tele Control System basically consists of a Tunnel Control Centre (TCC) interconnected to Instrumented Vehicles by a Public Mobile Network (PMN), that supports both Global System Mobile (GSM) connections and the General Packet Radio Service (GPRS) connections. The Tele Control System is designed to implement preventive safety actions in different tunnel scenarios (normal vehicular traffic, incidents, diffusion of emergency information) and the PMN is dimensioned for the expected throughput of voice and data, between the Instrumented Vehicles and the TCC, under such scenarios. The SAFETUNNEL Project designs the Tele Control System and implements a system Demonstrator, that is a prototypal subset of the Tele Control System . The validation of Tele Control System will be performed both by experimental tests and by modelling. A limited number of experimental tests are planned on the actual system Demonstrator; moreover a set of validation measures have to be predicted by system models, because the Demonstrator is not suitable for such measures. In fact the Demonstrator, that operates inside the tunnel, is not suitable for measures which would require long observation time inside the tunnel (that should be closed to the ordinary vehicular traffic, with loss of availability and money) and measures which would require irreproducible tunnel scenarios (i.e occurrence of incidents and emergency scenarios). Less than ever, the System Demonstrator is suitable for performance and availability measures, which are typically predicted by modelling and simulation and rarely performed by using experimental data from long, inadequate and costly observations of the whole system (and not of a part of it, that is the System Demonstrator). Due to the complexity of the Tele Control System and according to the Validation Plan, the system validation by modelling will not be exhaustive but will be focused on system relevant properties, that could affect the Tele Control System safety and timeliness [2], [3]. Validation by modelling will address relevant parts of the Tele Control System, including the PMN, which represents the most innovative and challenging research aspect of system. The present paper just deals with denial of service measures of the PMN [4], intended as performance measures explicitly tied to service degradation/recovery due to components failure and

repair activities (availability measures). Denial of service measures here are computed by considering both performance and availability measures, because performance measures, which ignore failures and recovery activities, but just consider resource contention, generally over estimate the system's ability to perform. On the other hand pure availability measures, where performance are not taken into account, tends to be too conservative. To compute denial of service measures (in terms of voice blocking probability and packet loss probability), we built modular sub models, hierarchically composed, by using Stochastic Activity Networks (SAN). At the first layer, we have built three sub models to compute pure unavailability and pure performance measures. Then, at the second layer, we have built two composed models , respectively for voice and data packet services. Each composed model joins the pure availability sub model and the related pure performance sub model, in order to compute the denial of service measure. The paper is organized as follows. Section 2 and 3 describe the basic elements of the Tele Control system and of the GSM/GPRS architecture. Section 4 deals with the PMN modelling assumptions and introduces the denial of service measures. Section 5 describes the modelling formalism: the Stochastic Activity Networks. Sections 6, 7, 8 describe the PMN models and measures. Some numerical results are reported in section 9. In section 10 there are some discussions and conclusions.

Tele Control System

The Tele Control System implements its safety functions[1], transferring voice, commands and data between Instrumented Vehicles and the Tunnel Control Centre. TCC must be able to exchange information with more than one Vehicle at the same time in bi-directional way. Particularly, informative messages are transmitted in uplink (from Vehicles on-board system to TCC) for the purpose of diagnosis and prognostics of vehicles. Commands/messages are transmitted in downlink (from TCC to a single vehicle or to a set of vehicles) for notification of a dangerous conditions inside the tunnel, or for setting/updating vehicle parameters (such as vehicle speed, safety intra-vehicles distance). For each Vehicle entering the Safe Tunnel monitored area, the TCC sets up a dedicated GPRS connection. TCP transport protocol is used to guarantee the correctness of data by means of integrity checks in the receiver and foreseeing a retransmission mechanism for bad-received packets. Each Vehicle is characterized by a TCP address (IP address + TCP port) in order to be able to communicate to the TCC that is provided of an analogous address too. Moreover, bidirectional voice calls, supported by GSM connection, are also provided between Vehicles and TCC, in case GPRS data transfer are not sufficient to manage an emergency.

GSM/GPRS architecture

GSM [5],[6] is a circuit-switched connection, with reserved bandwidth. At air interface, a complete traffic channel is allocated to a single Mobile Station (MS) for the entire call duration. A cell is formed by the radio area coverage of a Base Transceiver Station (BTS). One or more BTS are controlled by one Base Station Controller (BSC). Such a set of Stations form the Base Station Subsystem (BSS). A BSS can be viewed as a router connecting the wireless cellular network to the wired part of the network. GSM uses a mixed multiple access technique to the radio resources: Frequency Division Multiple Access/Time Division Multiple Access (FDMA/TDMA). Within each BSS, one or more carrier frequencies (FDMA) are activated, and over each carrier a TDMA frame is defined. TDMA allows the use of the same carrier to serve multiple MS. In the GSM system the frame is constituted by eight timeslots and so the same radio frequency can serve up to eight MS. A circuit (a channel) is defined by a slot position in the TDMA frame and by a carrier frequency. Typically one channel (time slot) is reserved to signaling and control. A MS can roam from a cell to a neighboring cell during active voice calls. Such a MS, that has established a voice call, and roams from a cell to another, must execute a handoff procedure, transferring the call from the channel in the old cell to a channel in the new cell entered by the MS. GPRS is a packet switched connection with shared, unreserved bandwidth. For data services, which is a bursty traffic, the use of GSM results in a highly inefficient resources utilization. For bursty traffic, a packet switched bearer service, such as GPRS, results in a much better utilization of the traffic channels. A radio channel will only be allocated when needed and will be released immediately, after the transmission of packets. With this principle more than one MS can share one physical channel (statistical multiplexing). In order to integrate GPRS services into the existing GSM architecture, a

[1] The Tele Control System safety functions include: 1)Vehicle Prognostics, 2)Access & Vehicle Control, 3)Vehicle Speed and Intra-Vehicles Distances Control, 4)Dissemination of Emergency Information.

new class of network nodes, called GPRS support nodes (GSN), are used. GSNs are responsible for the delivery and routing of data packets between the MS and the external packet data networks. A serving GPRS support node (SGSN) is responsible for the delivery of data packets from and to the MS [5]. GPRS exploits the same radio resources used by GSM. To cross the wireless link the data packets are fragmented in radio blocks, that are transmitted in 4 slots in identical position within consecutive GSM frames over the same carrier frequency [7].

Depending upon the length of the data packets, the number of radio blocks necessary for the transfer may vary. Mobile Stations execute packet sessions which are alternating sequences of packet calls and reading times. One time slot constitutes a channel of GPRS traffic, called Packet Data Traffic Channel (PDTCH) [5]. On each PDCH, different data packets can be allocated in the same TDMA frame or in different TDMA frames. When a user needs to transmit, it has to send a channel request to the network through a Random Access Procedure, which may cause collisions among requests of different users. In this case a transmission is tried. The number of maximum retransmissions is one of the GPRS access control parameters. Typically, one of the channels, randomly selected out of the available channels, is dedicated to GSM and GPRS signalling and control.

PMN modelling assumptions and measures

The dimensioning of the PMN accounts for several aspects including the length of the tunnel and the length of the tunnel monitored area, the recommended speed of vehicles and the safety distance between vehicles inside the tunnel, the number of carriage ways, the average and the worst demands of voice and data connections, the GPRS expected throughput per physical channel, the bit rate for the information exchange of each vehicle and the GSM expected connections. For GSM connection the same carrier frequency can serve up to eight vehicles. For GPRS connection, we assume that up to two vehicles are allocated in the same time slot, so the same carrier frequency can serve up to sixteen vehicles. One time slot (physical channel) is reserved as long as a voice call remains active, that is until the voice call is voluntarily released, then voice call generates an ON/OFF traffic on PMN. On the other hand, data transfer generates a bursty traffic (namely at vehicle registration/deregistration phases, in case of rare vehicle anomalies or incidents). One of the channels, randomly selected out of the available channels, is dedicated to GSM and GPRS signalling and control. Then the total number of available physical channels of our PMN is obtained from the product of the number of carriers per the number of channels per each carriers minus one, which represents the control channel. The PMN under analysis consists of one Base Station System (BSS), which contemporarily implements GSM and GPRS connections. Figure 1 shows the BSS with its essential components. GPRS connection is an updating service of the GSM architecture, which is born to deliver voice calls. We assume that GSM voice calls have higher priority than GPRS data transfer. That is, voice calls are set up as long as at least one physical channel is available in the BSS of interest; data packets can be transmitted only over the channels which are not used by voice connections. The handoff procedure [3], that allows roaming from a cell to a neighbouring cell is meaningful for GSM connections. Vice versa the handoff procedure is neglected for GPRS connections, since the duration of data transfer is typically much smaller than the time spent by a vehicle in a cell.

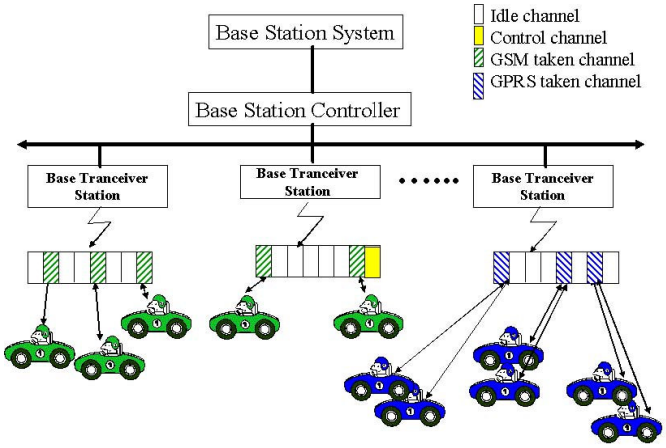


Figure 1 - PMN under analysis

To sum up, for the sake of building manageable models of our PMN, the following assumptions have been made: -we will focalize on a single Base Station System, constituted by one Base Station Controller and multiple Base Transceiver Stations -data exploits the same physical channels used by voice -channel allocation policy is priority of voice on data -we account for handoff procedure for voice connection - we neglect the possibility of the handoff procedure for data connection -one Control Channel (CCH) is dedicated to GSM and GPRS signalling and control; CCH is randomly assigned to a BTS -GPRS implements a point to point connection -each Instrumented Vehicle embeds a Mobile Station, which allows the contemporarily use of GSM and GPRS connections.

Denial of service measures: considering the PMN under analysis limited to one BSS, as showed in figure 1, the GSM and the GPRS services can be denied, due to at least one of the following contributes: a) the BSS, as a whole, becomes unavailable or b) the BSS is available and all its channels are full or c) the BSS is not completely available and all the channels in it, which are available, are also full. We named TSB, the Total Service Blocking Probability, as the denial of service measures of GSM and GPRS connections, due to the occurrence of at least one of the contributes a), b), or c). Regarding the contribute a) the fact that the BSS and its channels are unavailable, depends upon the failure/repair activities of BSS physical components, which include the Mobile Stations, embedded inside the Instrumented Vehicles, the Base Transceiver Stations and the Base Station Controller. BSS components are assumed to fail and be repaired with their own and independent rates. Actually, the reliability figures of Mobile Stations are significantly better than those of the other network comp onents, then we assume the MS as fault free. Each BTS can hosts eight traffic channels or, randomly, could hosts the Control Channel (CCH) plus seven traffic channels. To sum up, the BSS Total Unavailability (TU) is approximately:

$$TU = BCF + CCF + ATF \quad (1)$$

where:

-BCF is the unavailability of the Base Station Controller -CCF is the unavailability of the Control Channel (CCF) which depends upon the unavailability of the BTS which randomly can host it.

-ATF is the unavailability of all the BTS. When a BTS which does not host the CCH fails, its physical channels became unavailable and the BSS works in degraded conditions. If the failure of all the BTS (ATF) occurs, the consequence is still TU, the Total Unavailability of the BSS.

To compute the Total Blocking Probability (TSB) of our PMN, we have built modular stochastic models, hierarchically composed, by using Stochastic Activity Networks (SAN). Two different layers of modelling have been implemented. At the first layer, we built a model to compute the pure GSM/GPRS unavailability, TU, according to formula 1, which represents the contribute a) to TSB. At the same layer, we still built two separate models to respectively compute voice and data packet performances. Due to the assumption of priority of voice on data, the performance model of voice just takes into account the GSM connection, while the performance model of data packets has to take into account the contention of

the same physical channels between GSM and GPRS. The performance models compute the probability of having all available channels full and represent the contribute b) to TSB. Then, at the second layer of modelling, we have built two composed models. A composed model joins the pure availability model and the voice performance model to compute the whole TSB for voice connections. The other composed model joins the pure availability model and the data packet performance model, to compute the whole TSB for data packet connections. We have to consider that TSB completely measures the loss of voice for GSM connection, because voice is not retransmitted. On the other hand, for GPRS connection, TSB affects the loss of data, but does not directly measure it. In fact data packets can be accumulated into a queue and retransmitted according to GPRS access control parameters [5]. Then, for GPRS connection, other than TSB, we also compute the probability of data packet loss for exceeding the buffer capacity and the probability of data packet loss for exceeding the maximum number of data packet sessions which can be simultaneously opened.

Stochastic Activity Networks

Stochastic Activity Networks (SAN) are a modelling formalism which extends Petri Nets [8]. The basic elements of SAN are places, activities, input gates and output gates. Places in SAN have the same role and meaning of places of Petri Nets. They contain an arbitrary number of tokens. Activities are equivalent to transitions in Petri Nets. They can take a certain amount of time to be completed (timed activities) or no time (instantaneous activities). Each activity may have one or more input arcs, coming from its input places (which precedes the activity) and one or more output arcs going to its output places (which follow the activity). In absence of input gate and output gate, the presence of at least one token in each input place makes it able to fire and after firing one token is placed in each output place. Input gates and output gates, typical constructs of SAN, can modify such a rule, making the SAN formalism more rich to represent actual situations. Particularly, they consist in predicates and functions, written in C language, which contain the rules of firing of the activities and how to distribute the tokens after the activities have fired. As in Petri Nets, a marking depicts a state of the net, which is characterised by an assignment of tokens to all the places of the net. With respect to a given initial marking, the reach ability set is defined as the set of all markings that are reachable through any possible firing sequences of activities, starting from the initial marking. Other than the input and output gates, which allow to specifically control the net execution, SAN offers two more relevant high-level constructs for building hierarchical models: REP and JOIN. Particularly. such constructs allow to build composed models based on simpler sub-models, which can be developed independently and then replied and joined with others sub-models and then executed. The SAN model specification and elaboration is supported by Möbius tool, developed by University of Illinois. The tool allows to specify the graphical model, to define the performance measures through reward variables, to compute the measures by choosing a specific solver to generate the solution.

The availability sub model

To compute TU (formula (1)), we have built the availability sub model of figure 2. The sub model includes the failure/repair behaviour of the Base Station Controller and the failure/repair behaviour of all the controlled Base Transceiver Stations, according to the terms of formula (1). A failed BTS hosts the Control Channel (CCH) with probability c, or complementary host the CCH, with probability 1 – c. If the failed BTS hosts the CCH, the BTS failure implies the failure of the Control Channel, and in turn, the failure of the whole PMN. If the BTS, doesn't host the CCH, the BTS failure just implies the loss of the physical channels supported by it (eight channels/timeslots).

The marking of place BTS_UP represents the number of Base Transceiver Stations which are not failed. The firing of the activity BTS_Fail represents the failure of the BTS component. If the failed BTS hosts the CCH, it makes the whole BSS down (output gate TU_CCH, shown in table 1). If the failed BTS doesn't host the CCH, the channels which are currently up are decremented by the number of channels associated to the failed BTS (output gate BTS_loss). The marking of the place BTS_DOWN represents the number of failed BTS; one token in the place CCH_DOWN represents the CCH failure. The firing of the activities BTS_Repair and CCH_Repair represents the repair activities of the related BTS component. One token in place BCS_UP represents that the BCS is not failed. One token in place BCS_DOWN, consequent to the firing of the activity BCS_Fail, represents the BCS failure. On the failure of the BCS, the whole BSS goes down and all the channels are lost (output gate TU_BCS) The marking of the place working_channels represents the number of available and idle channels. The

marking of the place channels_in_service represents the number of available and connected channels. After the repair activities (CCH_repair, BCS_repair, BTS_repair) the channels are again up and ready to be taken in service (ouput gates BTS_ON, BCS_ON, CCH_ON). The firing time of the activities is assumed to follow a negative exponential distribution.

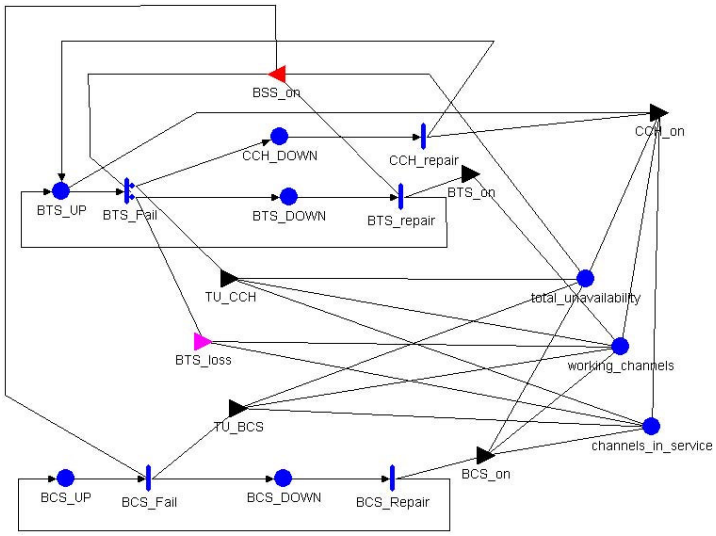


Figure 2 -The availability sub model

Output Gate Attributes: TU_CCH
Field Name

Field Value total_unavailability->Mark()=1; Function

working_channels->Mark()=0; channels_in_service->Mark()=0;

Output Gate Attributes: TU_CCH	
Field Name	Field Value

Function	total_unavailability->Mark()=1;
	working_channels ->Mark()=0;
	channels_in_service->Mark()=0;

Table 1 -Definition of the output gate TU_CCH

GSM denial of service composed model

The GSM denial of service composed model computes the Total Blocking Probability (TSB) for voice service. To compute TSB, we consider our PMN as completely dedicated to the GSM services, due to the assumption of the priority of voice on data. The GSM denial of service composed model takes into account the contention of the radio channels from the voice calls (either new or continuous) modelled by a pure performance sub model, combined with the possible loss/recovery of the radio channels due to the failure/repair activity of the BSS components. Particularly, GSM denial of service composed model, figure 3, has been built joining the availability sub model of section 6, and the GSM Performance sub model, which models the pure performance aspects of the GSM service.

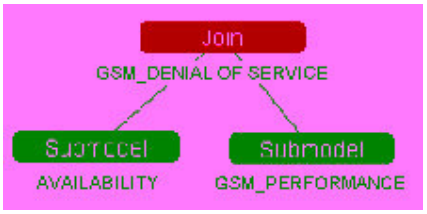


Figure 3 - GSM denial of service composed model

GSM performance sub model: the GSM performance sub model, figure 4, computes two performance measures: the New Call Blocking probability and the Continuous (handoff) Call Blocking probability, due to all N channels full and not failed. It is assumed that blocked calls are lost and not reattempted. The GSM performance sub model represents the PMN with a number of servers which represents the number of available channels. Moreover, a limited number of available channels, named guard channels, are exclusively reserved for the handoff calls. Referring to figure 4, the marking of the place working_channels represents the number of not-failed channels, that are currently idle. The marking of the place channels_in_service represents the number of not-failed channels, that are currently busy. The firing of transition T_new_call represents the arrival of new calls and the firing of transition T_continuous_call represents the arrival of a handoff call from neighbour cells. A handoff call will be dropped only when all channels are busy. This is realised by the input gate I_Total_channels which enables the transition T_continuous_call to fire when all not-failed channels are busy. A new call will be blocked if there are no more than the number of the reserved channels for handoff calls. This is realised by the input gate Reserved_channels, which enables the transition T_new_call to fire when all not-failed and not reserved channels are busy. The firing of the transitions T__call_completation and T_handoff_out respectively represent the completion of a call and the departure of an outgoing handoff call. All activities are assumed exponentially distributed.

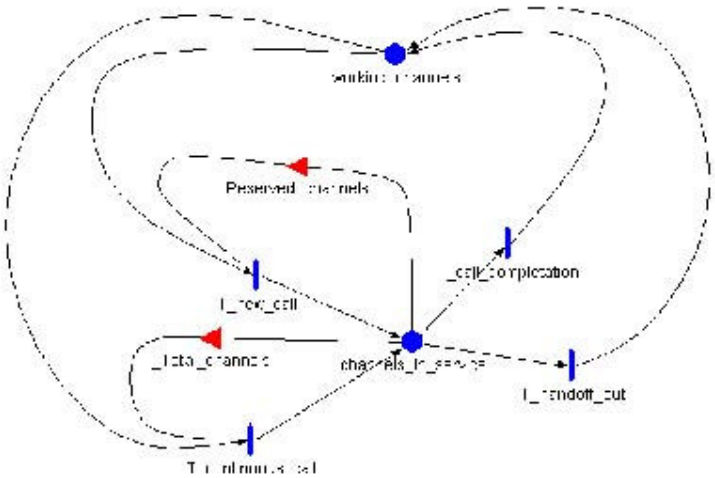


Figure 4 - The GSM performance sub model

GSM&GPRS denial of service composed model

The GSM&GPRS denial of service composed model (figure 5) computes the Total Blocking Probability (TSB) on packet data service. The composed model joins the GSM&GPRS performance sub model, that represents the contention of the radio channels from the voice calls and data packets transfer request and the availability sub model that represents the possible loss/recovery of the radio channels due to the failure/repair activity of the BSS components.

In case of GPRS, TSB does not directly measure the loss of information contained in data packets because they can be accumulated into a queue and retransmitted. Then, for GPRS connection, other than TSB, we also compute the probability of data packet loss for exceeding the buffer capacity and the probability of data packet loss for exceeding the maximum number of data packet sessions which can be

simultaneously opened.

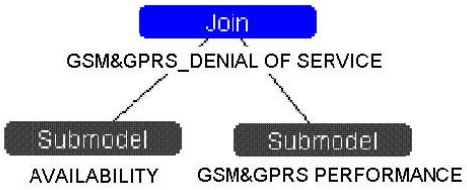


Figure 5 - The GSM&GPRS denial of service comp osed model

GSM&GPRS performance sub model: the GSM&GPRS performance sub model computes the pure performance aspects of the GPRS service, which contends physical channels to the GSM service. Voice calls are set up as long as at least one channel is available in the PMN, while data packets can be transmitted only over the channels which are not used for voice service. A vehicle, which needs to communicate with Tunnel Control Centre or vice versa, tries to open a packet session. If the current number of open data packet sessions is less than the maximum number of data packet sessions which can remain simultaneously active, then a new data packet session can be opened. Into an active data packet session, the incoming data packets are queued in a buffer, as a sequence of radio blocks. Once in the buffer, the radio blocks can be transmitted with the proper GPRS transmission rate. The transfer of radio blocks over the radio link can be either successful, thus allowing the removal of the radio block from the buffer, or results in a failure; in the last case, the radio block is retransmitted.

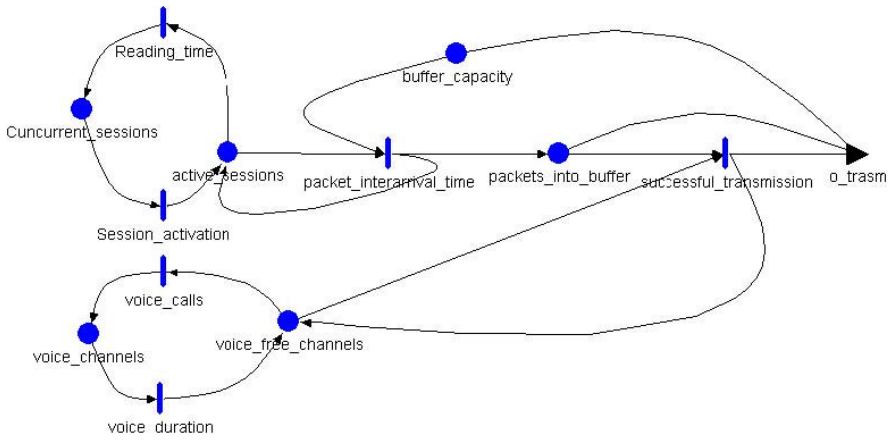


Figure 6 - The GSM&GPRS performance sub model

Referring to figure 6, if at least one token is in place concurrent_section a data packet session is opened, by the firing of session_activation activity. As a consequence one token is added in place active_section. Named D, the number of maximum simultaneously active data packet sessions and named d, the number of currently opened data packet sessions, a new session can be opened at the condition that $d < D$. Inside an open data packet session, data packets arrive with the rate of packet_interarrival_time activity and are queued into packet_into_buffer place. As a first step, we assume that one data packet has the length of one radio block, so each data packet increments the buffer by one unit (one radio block) at the condition that the buffer is not full (if $b < B$, where b is the current values of the radio blocks in the buffer and B is the buffer capacity). Such a condition is controlled by the marking of buffer capacity place. The radio blocks queued in the buffer are transmitted during the same set of 4 TDMA frames by the successful_transmission activity which keep into account that the radio block that can be served by the currently available channels (the ones not being occupied by voice).

Some numerical results

We conduct availability, performance and denial of service measures on voice and data services, executing the models described in the previous sections, by Mobius analytical solver [8]. The input parameters and their numerical values are summarized in Table 2, 3 and 4.

Parameter	Value
Rate of BSC_fail	2,31 E-4 h-1
rate of BSC_repair	1 h-1
Rate of CCF_fail	3.47 E-4 h-1
rate of CCF_repair	0,5 h-1
Rate of BTS_fail	3.47 E-4 h-1
rate of BTS_repair	0,5 h-1
Number of BSC	1
Number of BTS	4
n. of channels of a BTS	8
Number of CCH	1

Table 2 - Input parameters and values of the availability sub model

Parameter	value
arrival rate of new calls	0,27 s-1
duration of the calls	180 s
arrival rate of handoff calls	0,027 s-1
duration of outgoing handoff calls	80 s

Table 3 - Input parameters and values of the GSM performance sub model

Parameter	Value
arrival rate of voice calls	0,5…2,5 s-1
duration of voice calls	180 s
rate of session activation	2 s-1
session reading time	15 s
Packets inter arrival rate	0,0242 s1
rate of suc. packet transmission	0,0513 s1
buffer capacity (B)	100
n. of max opened sessions (D)	10,30,50

Table 4 - Input parameters and values of the GSM&GPRS performance sub model

Some numerical results are shown in figure 7 and 8. Figure 7 shows the Total Service Blocking Probability

(TSB) for voice service, versus time, computed by the GSM denial of service composed model. The computation of TSB is performed by using the total_blocking reward variable, which increments its value of 1 when the number of available channels, ready to serve, becomes equal to zero.

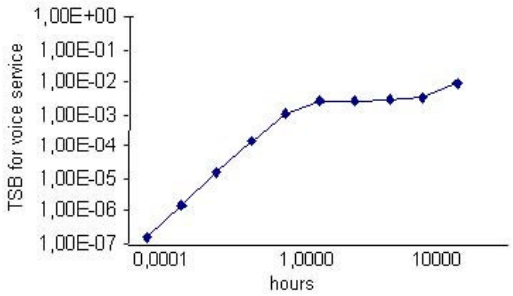


Figure 7 - Total Service Blocking (TSB) probability for voice service

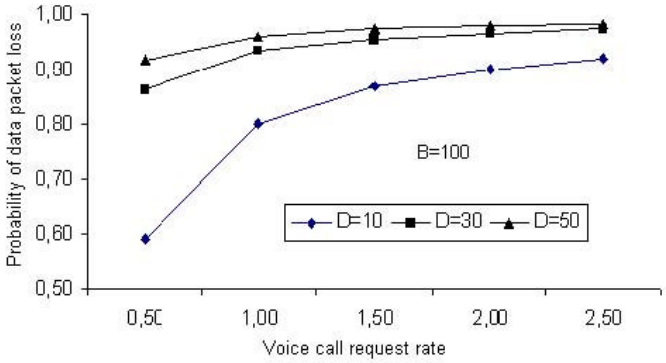


Figure 8 -Probability of data packets loss Figure 8 shows the probability of data packets loss for data service, due to the buffer overload, versus voice call request rate, computed by the GSM&GPRS performance sub model. The measures have been computed for different values of the maximum number of simultaneously opened data packet sessions (D=10, 30,50). We assume buffer capacity, B=100.

Conclusions and future research

The work presented in this paper is in the framework of validation by modelling of a Tele Control system, based on a Public Mobile Network (PMN). We have computed measures of the denial of service for GSM and GPRS connections, such as the Total Service Blocking Probability (TSB), to better understand the effects of the degradation of the performance and of the availability of the PMN on the Tele Control system main functions. We have built modular sub models, hierarchically composed, by using Stochastic Activity Networks. Two different layers of modelling have been implemented. At the first layer, we built separate sub models to compute the pure unavailability and the pure performance for voice and data packet services. At the second layer of modelling, we have built two composed models joining the availability sub model and the performance sub models. The first numerical results have been presented. Currently, we are extending our research to include the impact of security crushing in a global dependability model for communication networks.

Acknowledgements – The research work presented in this paper has been partially supported by the IST – 1999-28099 SAFETUNNEL project and its consortium: Centro Ricerche Fiat (I), Renault VI (F), TILAB (I), SITAF (I), SFTRF (F), Fiat Engineering (I), TÜV (D), Un. Ben Gurion (Isr), Enea (I)

References

1 Project IST – 1999 – 28099, SAFETUNNEL -http://www.crfproject-eu.org

2 E. Ciancamerla, M. Minichino, S. Serro, E. Tronci -Automatic Timeliness Verification of a Public

Mobile Network – Safecomp 2003, 22nd International Conference on Computer Safety, Reliability and Security – Edinburgh, UK - September 23-26, 2003
3 E. Ciancamerla, M. Minichino, S. Serro, E. Tronci - Worst Case Time Analysis of a Wireless Network via Model Checking - International Conference "Automation within new global scenarios", 30th BIAS edition - Milan 19, 20, 21 November 2002
4 K. S. Trivedi, Xiaomin Ma – Performability Analysis of Wireless Cellular Networks – SPECTS2002 and SCSC2002 – July 2002
5 ETSI – Digital Cellular Telecommunication System (Phase 2+); General Packet Radio Service (GPRS) - GSM 04.60 version 8.3.0
6 C. Bettsletter, H. Vogel, J. Eberspacher - GSM phase 2+ - General packet radio service GPRS: architecture, protocols and air interface – IEEE Communications Survey vol. 2 , n.3 – 1999
7 M. Meo, M. Ajmone Marsan, C. Batetta – Resource Management Policies in GPRS Wireless Internet Access System – 2002 IEEE
8 W.H. Sanders, W.D. Obal, M.A.Qureshi, F.K. Widjanarko – The UltraSAN modelling Environment Performance Evaluation J. special issue on performance modelling tools, vol. 24, pp 89 - 115, 1995

EUROCONTROL - Systemic Occurrence Analysis Methodology (SOAM)

Tony Licu (1), Brent Haywar (2), Andrew Lowe (3)

1. Eurocontrol (antonio.licu@eurocontrol.int, www.eurocontrol.int/ssap),
2. (bhayward@dedale.net, www.dedale.net) and
3. (alowe@dedale.net) – Dédale Asia Pacific

Abstract
The Safety Occurrence Analysis Methodology (SOAM) developed for EUROCONTROL is one of a number of accident investigation methodologies based on the Reason Model of organisational accidents. The purpose of a systemic occurrence analysis methodology is to broaden the focus of an investigation from the human involvement[1] to include analysis of the latent conditions deeper within the organisation that set the context for the event. Such an approach is consistent with the tenets of Just Culture[2] in which people are encouraged to provide full and open information about how incidents occurred, and are not penalised for errors.

A truly systemic approach is not simply a means of transferring responsibility for a safety occurrence from front-line employees to senior managers. A consistent philosophy must be applied, where the investigation process seeks to correct deficiencies wherever they may be found, without attempting to apportion blame.

Keywords
Just Culture, SOAM, safety occurrences, safety analysis, SHELL model, barriers, defences, Strategic Safety Action Plan (SSAP), EUROCONTROL,

Introduction
As a direct result of the runway incursion accident at Milan Linate airport (Italy in October 2001) and the mid-air collision near Überlingen (Germany in July 2002), EUROCONTROL established a High Level European Action Group for ATM Safety (AGAS) to examine existing procedures and standards. The objective was to propose enhancements in ATM safety within the 41 States of the European Civil Aviation Conference (ECAC).

By gathering together experienced safety experts from across the industry to scrutinise all aspects of ATM safety, AGAS was able to identify the areas where most benefit will be gained by improving safety in the short term. As a consequence a Strategic Safety Action Plan has been structured to provide quick implementation solutions for improving Air Traffic Management safety throughout Eight High Priority Action Areas:

1. Safety Related Human Resources in ATM
2. Incident Reporting and Data Sharing
3. Airborne Collision Avoidance System
4. Ground-Based Safety nets
5. Runways and Runways Safety
6. Enforcement of ESARRs and monitoring of their implementation
7. Awareness of Safety Matters
8. Safety and Human Factors Research and Development

---

1 Also known as "active failures of operational personnel" under original Reason model

2 For further detail on Just Culture, see: EUROCONTROL. (2004). *EAM2/GUI6: Establishment of "Just Culture" Principles in ATM Safety Data Reporting* (Edition 0.1 25 November 2004). Brussels: Author.

SOAM has been developed to support the objectives of AGAS Priority Area 2, Incident Reporting and Data Sharing, by:

- Providing an investigation methodology that can be applied locally by a large number of trained users, across a wide variety of occurrences. Occurrence data collection would then be a dispersed rather than centralised and specialised activity, increasing the potential quantity of data analysed;
- Establishing a dedicated investigation terminology, providing a common language for trained users that facilitates data exchange and understanding;
- Supporting Just Culture principles, which are closely aligned with the philosophy underlying the investigation technique. A comprehensive training program to roll-out the new process would incorporate awareness and education on the benefits of a Just Culture and of open reporting;
- Providing standardised principles for Air Navigation Service Providers (ANSPs), investigators and airspace users on generating valid, effective remedial actions once contributing factors are identified; and
- Providing additional structure and focus to the common taxonomy for reporting and investigating ATM safety occurrences.

Most importantly, SOAM will support one of the most critical *Harmonisation*[3] objectives, by providing a common methodology for the identification of causal factors across the aviation industry. This has the potential to enhance *Data Sharing and Lesson Dissemination* by:

- Providing a simple framework (based on principles drawn from the now widely-disseminated and recognised Reason Model) for sharing safety information, covering in particular the contributing factors and remedial actions;
- Standardising the way safety improvement actions are generated; and
- Making it simpler to summarise the outcome of real investigated occurrences for publication, for example in issues of Safety News.

SOAM Approach

The investigation philosophy on which the SOAM approach is based is adapted from ICAO Annex 13, as follows:

"The fundamental purpose of safety investigation is the prevention of further occurrences. It is not our task to apportion blame or liability" [4]

Safety occurrences are by definition events in which there was a deviation from the desired system state, resulting in loss or damage to equipment or personnel, or increased potential for such outcomes. Every occurrence provides an opportunity to study how the deviation occurred, and to identify ways of preventing it from happening again.

The objectives of safety occurrence investigation are to:

- Establish what happened
- Identify local conditions and organisational factors that contributed to the occurrence
- Review the adequacy of existing system controls and barriers
- Formulate recommendations for corrective actions to reduce risk and prevent recurrence
- Identify and distribute any key lessons from the safety occurrence
- Detect trends that may highlight specific system deficiencies or recurring problems

[3] EUROCONTROL. (2003). EAM2/GUI5: Harmonisation of Safety Occurrence Severity and Risk Assessment. (Edition 0.1, 05 June 2003). Brussels: Author.

[4] International Civil Aviation Organization. (1994). Annex 13 to the Convention on International Civil Aviation: Aircraft accident and incident investigation, Eighth edition, July 1994. Montreal: Author.

SOAM Process Overview
The SOAM process follows the sequence depicted below:

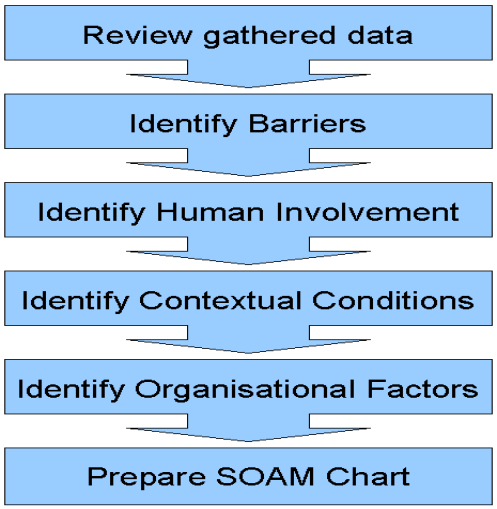


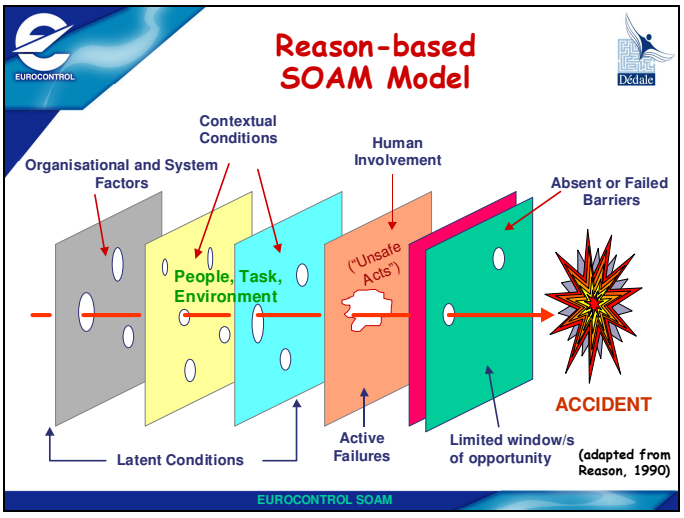
Figure 1 – The SOAM Process

Systemic Occurrence Analysis Method
The Systemic Occurrence Analysis Method (SOAM) is one of several accident analysis tools based on principles of the well-known "Reason Model" of organisational accidents (Reason, 1990, 1991).

SOAM is a process for conducting a systemic analysis of the data collected in a safety occurrence investigation, and for summarising this information using a structured framework and standard terminology. As with some root-cause analysis investigation methods, SOAM draws on the theoretical concepts inherent in the Reason Model, but also provides a practical tool for analysing and depicting the inter-relationships between all contributing factors in a safety occurrence.

SOAM allows the investigator to overcome one of the key historical limitations of safety investigation – the tendency to focus primarily on identifying the errors – those intentional or unintentional acts committed by operators – that lead to a safety occurrence. This so-called ‘person approach’ to accident investigation can only provide a superficial explanation of an occurrence because it does not consider the underlying root causes which may have contributed to, or allowed, the individual actions which triggered the event. The person approach considers only the transparent ‘active failures’ or unsafe acts, rather than searching for the less obvious contributing factors or ‘latent conditions’ within the system.

Reason's original model has been adapted and refined within SOAM. The nomenclature has been altered in accordance with a "Just Culture" philosophy, reducing the implication of culpability and blame by both individuals and organisations. In SOAM, 'Unsafe Acts' are referred to as *Human Involvement*, 'Psychological Precursors of Unsafe Acts' as *Contextual Conditions*, and 'Fallible Decisions' as *Organisational and System Factors*. The SOAM version of the Reason Model is shown below.



Like other systemic analysis techniques, SOAM forces the investigation to go deeper than a factual report that simply answers questions such as “What happened, where and when?" First, data must be collected about the conditions that existed at the time of the occurrence which influenced the actions of the individuals involved. These in turn must be explained by asking what part the organisation played in creating these conditions, or allowing them to exist, thereby increasing the likelihood of a safety occurrence. SOAM thus supports the fundamental purpose of a safety investigation - to understand the factors which contributed to an occurrence and to prevent it from happening again.

SOAM is aligned with and supports "Just Culture" principles by adopting a systemic approach which does not focus on individual error, either at the workplace or management level. It avoids attributing blame by:

- Removing the focus from people’s actions, instead seeking explanation for the conditions that shaped their behaviour; and
- Identifying latent organisational factors that allowed less than ideal conditions to exist, under which a safety occurrence could be triggered.

As with the original Reason Model, SOAM can be applied both reactively and proactively.

The process can be applied to any new occurrence, and is also suitable for the retrospective analysis of previously investigated occurrences in an attempt to extract additional learning for the promotion of safety. SOAM can also be applied proactively to generic occurrences (e.g., level busts, separation minima infringements, runway incursions, etc.) or hypothetical events. These applications result in a comprehensive analysis of the absent or failed barriers and latent conditions that are commonly found to contribute to such events, thereby identifying areas of organisational weakness that need to be strengthened to improve safety and prevent future occurrences.

Gathering factual data

While there is no definitive or prescribed method for the gathering of investigation data, it is useful to gather data within some form of broad descriptive framework, to help with the initial sorting of facts. The SHEL Model (Edwards, 1972) provides the basis for such a descriptive framework. An adaptation of the SHEL Model is depicted in figure 3.

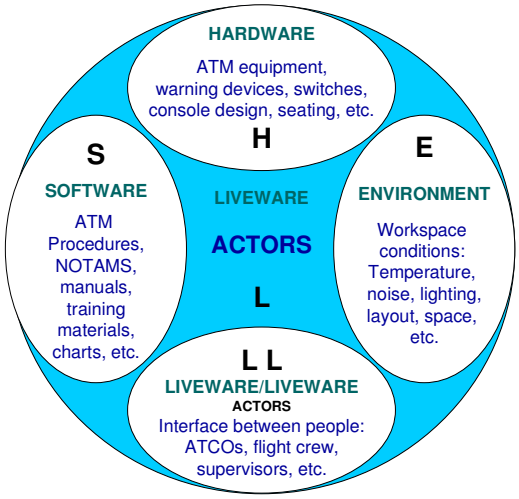


Figure 3 – The modified SHEL model

Data should be gathered across five areas (the four original areas of the SHEL model, and an extra fifth element – organisation):

- Liveware – the human element (personnel)
- Software – procedures, manuals, symbology, etc.
- Hardware – equipment, workplace layout, etc.
- Environment – workspace conditions, noise, temperature, or other factors that affect human operators
- Organisation - organisational decisions/actions that impact on people in the workplace.

While the data gathering and analysis phases in an investigation are typically depicted as discrete, in reality they are part of a recursive process. After an initial data collection phase, a preliminary analysis can be conducted, which will identify gaps that can be filled by further data gathering. This process will continue until the systemic analysis has eliminated unanswered questions and reached a logical conclusion.
Examples of the types of data which can be collected under each SHEL element are to be found in the detailed guidelines on SOAM (EAM2-GUI8).

SOAM Analysis
Having collected the data, the first stage of the SOAM analysis involves sorting each piece of factual information into an appropriate classification. This is a progressive sorting activity which can be conducted as a group exercise if the investigation is being conducted by a team. Each fact is dealt with in turn, and subjected to two tests:

TEST 1: Does the fact represent a condition or event that contributed to the eventual occurrence?
Test 1 ensures that information that did not contribute to the occurrence is excluded from the SOAM analysis process. If the information is important, it can be detailed in a separate section of the investigation report.

TEST 2: Which one of the following categories can the fact be classified as:

- Absent or Failed Barrier
- Human Involvement
- Contextual Condition
- Organisational Factor

Check Questions are supplied for each SOAM category to assist with the classification of items under Test 2. At each stage of the SOAM process the relevant check question should be applied to ensure that the item being considered fits within the definition of the category for which it is being considered.

Identifying Absent/Failed Barriers
Complex socio-technical systems typically contain multiple barriers or defences to protect the system against hazards and undesired events. Barriers protect the system against both technical and human failures. Absent or failed barriers are the last minute measures which failed or were missing, and therefore did not (a) prevent an action from being carried out or an event from taking place; or (b) prevent or lessen the impact of the consequences.

A key objective of the investigation process is to identify barriers that failed to prevent the occurrence or minimise its consequences, or that could have prevented the occurrence had they been in place, and to recommend action to strengthen these.
The first step of the SOAM process involves identifying the barriers[5] which failed or were absent at the time of the occurrence. The following six barrier types (Awareness, Restriction, Detection, Control and Interim recovery, Protection and Containment and Search and Rescue) represent successive lines of defence, beginning with awareness and understanding of risks and hazards in the workplace. If this first line of defence is breached, subsequent barriers (restriction, detection, and so on) are designed to contain the situation and limit adverse consequences as control is progressively lost.

Identifying Human Involvement
Following identification of the relevant absent or failed barriers, the next step is to identify the contributing human actions or non-actions that immediately preceded the safety occurrence. The question at this stage should not be *why* people behaved as they did, but simply *what were their actions/inactions* just prior to the event.

SOAM analyses the human involvement in a safety occurrence using an existing model of information processing. The tasks performed by an Air Traffic Controller (ATCO) involve multiple forms of information processing, including accurate detection, integration and interpretation of information, as well as planning, projecting and decision making. An information processing model is thus a logical component of an ATM occurrence analysis methodology, enabling a comprehensive representation of the steps that might be performed by a controller as an occurrence unfolds. The information processing model selected for use with this methodology is Rasmussen's Decision Ladder technique (1982).[6] Like similar models, it assumes that information is processed in stages, beginning with the detection of information and ending with the execution of an action.

The analysis of human involvement using the Decision Ladder technique[7] provides the foundation for the next stage of the SOAM process which focuses on trying to understand why people acted as they did, through examination of the *contextual conditions* in place at the time of the occurrence.

Identifying Contextual Conditions
Contextual conditions describe the circumstances that exist at the time of the safety occurrence that can directly influence human performance in the workplace. These are the conditions that promote the occurrence of errors and violations.

In the occurrence investigation process, contextual conditions can be identified by asking “What were the conditions in place at the time of the safety occurrence that help explain why a person acted as they did?"

Five categories of contextual conditions can be distinguished, two relating to the local workplace, and three to people:

- Workplace conditions

[5] **NB:** While James Reason used the term 'Defences' in his modelling of organisational accidents, 'Barriers' is the terminology preferred in SOAM. For detailed discussion of the concept of barriers see Erik Hollnagel's work, in particular: Hollnagel, E. (2004). *Barriers and accident prevention.* Aldershot, UK: Ashgate

[6] Rasmussen, J. (1982). Human errors: a taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4,* 311-333.

[7] For the sake of keeping the paper brief the Decision Ladder technique is not reproduced.

- Organisational climate
- Attitudes and personality
- Human performance limitations
- Physiological and emotional factors

Identifying Organisational Factors (ORF)
ORFs describe circumstances which pre-existed the occurrence and produced or allowed the existence of contextual conditions, which in turn influenced the actions and/or inactions of staff. A total of 12 ORFs have been identified as frequently contributing to ATM safety occurrences. The factors and their corresponding two-letter codes are summarised in the table below.

Code	Organisational Factors
TR	Training
WM	Workforce Management
AC	Accountability
CO	Communication
OC	Organisational Culture
CG	Competing Goals
PP	Policies and Procedures
MM	Maintenance Management
EI	Equipment and Infrastructure
RM	Risk Management
CM	Change Management
EE	External Environment

Table 1 – Organisational Factors[8]

The SOAM Chart
The final product of the systemic occurrence analysis process is a summary chart depicting:

- The individual contributing factors – grouped according to the layers of the methodology as Barriers, Human Involvement, Contextual Conditions and Organisational Factors; and
- Horizontal links representing the association between a contributing factor at one level (eg., a human action), and its antecedent conditions (ie., the context in which the action took place).

Completing Links
In completing the links in the SOAM summary chart, facts at different levels should be linked if one is thought to have influenced the other. For example, if a contextual condition (e.g., fatigue) is considered to have influenced an action (e.g. delayed detection of conflict) then a linking line should be drawn between them. Similarly if an organisational factor (e.g., poor workforce management) is considered to have created a contextual condition (e.g., fatigue), or allowed it to continue to exist, then a link should be drawn between them.

An example SOAM chart is shown below in Figure 4.. In this example, data from the investigation of the October 2001 Milan runway collision has been employed to build a graphical representation of the circumstances surrounding the occurrence using the SOAM technique.

[8] Further information on each Organisational Factor, including illustrative case studies, is provided in *EAM2/GUI8: Guidelines on the Systemic Occurrence Analysis Methodology (SOAM).*

This end product of SOAM is very useful for briefing others and sharing lessons gained from identification of the circumstances surrounding an occurrence.

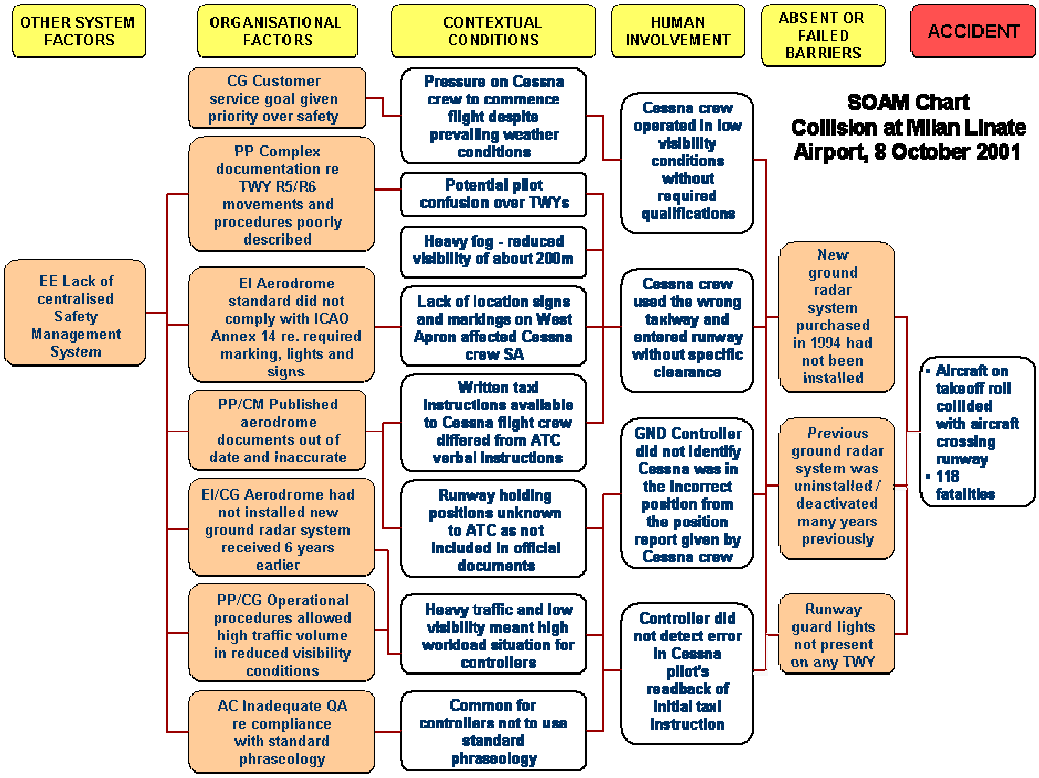


Figure 4 – Sample of SOAM Chart for Linate Accident

Formulating Recommendations

The formulation of recommendations for corrective action is a critical final element of the occurrence investigation process. The relevance, quality and practicality of remedial recommendations made following an investigation will determine their acceptability to those in a position to implement safety improvements.

This section describes the logical process within SOAM for generating recommendations that:

- Are directly and clearly linked to the SOAM analysis
- Are focussed on findings that are amenable to corrective action
- Reduce the likelihood of a re-occurrence of the event, and/ or reduce risk

In formulating recommendations, the SOAM process requires that the following two elements be addressed:

- The deficient Barriers (absent or failed), and
- The *Organisationa/Other System Factors*

See also Figure 4 above for exemplification.

Each failed or absent barrier must be addressed by at least one recommendation for corrective action. Each identified organisational factor must also be addressed by at least one recommendation, unless it is already adequately covered by a previous recommendation.

For example, a deficient warning system may be identified as a *failed barrier* as well as an *equipment and infrastructure* and/or *maintenance management factor* at the organisational level, but a single recommendation for corrective action may suffice.

Ensuring that recommendations correspond with the identified deficiencies in barriers and organisational factors will ensure that all latent conditions unearthed by the investigation analysis processes are addressed by recommended remedial action/s. This can also help to eliminate the problem of extraneous recommendations being made by exuberant investigators on matters of personal interest which were not identified as contributing factors in the occurrence at hand.

Summary

It is proposed in principle that the goal of improved system safety will be served by conducting some level of evaluation or investigation into all occurrences. This principle depends on the availability of a simple, systemic analysis methodology that can be applied reliably to all levels of occurrence. While highly competent investigators will always be required for complex, high level investigations, SOAM is suitable for use with all levels of occurrence, and is particularly suitable for use on lower level occurrences by investigators with relatively little training and experience.

Finally, SOAM encourages a "clinical" approach to the analysis of an event, seeing each investigation as a stand-alone, structured problem-solving activity. This contrasts with epidemiological approaches to safety prevention, that rely on checklists of "causal factors" and database analysis to target remedial actions. SOAM is designed to progressively develop a complete understanding about what happened, and address the latent conditions that will not only prevent a similar event, but strengthen the multiple layers of an organisation's operations that make it safe.

References

1. Edwards, E. (1972). Man and machine: Systems for safety. In *Proceedings of British Airline Pilots' Association Technical Symposium* (pp. 21-36). London: BALPA.

EUROCONTROL. (2003). *EAM2/GUI5: Harmonisation of Safety Occurrence Severity and Risk Assessment. (Edition 0.1, 05 June 2003).* Brussels: Author.

EUROCONTROL. (2003). *EAM2/GUI6: Establishment of "Just Culture" Principles in ATM Safety Data Reporting* (Edition 0.1 25 November 2004). Brussels: Author.

2. Hollnagel, E. (2004). *Barriers and accident prevention.* Aldershot, UK: Ashgate.

3. International Civil Aviation Organization. (2001). *Annex 13 to the Convention on International Civil Aviation: Aircraft accident and incident investigation, Ninth edition, July 2001.* Montreal: Author.

4. Rasmussen, J. (1982). Human errors: a taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4,* 311-333.

5. Reason, J. (1990). *Human error*. New York: Cambridge University Press.

6. Reason, J. (1991). Identifying the latent causes of aircraft accidents before and after the event. *Proceedings of the 22nd ISASI Annual Air Safety Seminar,* Canberra, Australia. Sterling, VA: ISASI.

Safeguarding information intensive critical infrastructures against novel types of emerging failures

C. Balducelli, S. Bologna, L. Lavalle, G. Vicoli

ENEA -Italian National Agency for new Technology, Energy and the Environment "Casaccia" Research Centre, Rome Email : claudio.balducelli@casaccia.enea.it

Abstract
The complexity of Information Intensive Critical Infrastructures, like electricity networks, telecommunication networks and public transportation networks is today augmented much more than in the past: such complexity augments the number of possible failures and anomalous working conditions and consequently decreases the survivability of the infrastructures. In this paper the possibility is investigated to detect early anomalies and failures inside information intensive critical infrastructures by the introduction of a population of agents being "self-aware" about the normal working conditions of the infrastructure itself. This new approach has the objective to improve the performance of the most popular signature based algorithms for intrusion detection, and makes use of different classes of time-oriented algorithms based on artificial intelligence paradigm. It has the advantage to work also in presence of unknown and unexpected types of attacks or failures. The results of the tests executed inside an emulated SCADA (Supervisory Control And Data Acquisition) system for electrical power transmission grid, and a proposal for the future integration inside real SCADA systems are also reported.

1. Introduction

Recognition of anomalies and failures, produced also by malicious attacks, inside large and complex systems, like the critical infrastructures on which all industrial countries depend on, seems not always feasible, applying predefined rules and procedures. When the interdependencies among the different components of a distributed system increase, the number of possible dangerous consequences diverges. Nowadays, more than in the past, many types of either accidental faults or deliberate attacks to complex infrastructures are *new*, unusual and different from the well known and experimented ones. Also the control and supervisory systems of the most critical infrastructures, as gas transport pipelines, oil refineries, power plants, water supplies systems and electrical power grids, generally named as SCADA (Supervisory Control And Data Acquisition) systems, are today more vulnerable[1]. The majority of SCADA and Digital Control Systems actually utilised by the energy utilities were developed many years ago, long before public and private networks or desktop computers became a common part of business operations. These kind of cyber-infrastructures worked with small degree of connectivity to the external networks and, for such reason, were more secure and less vulnerable respect to the modern ones. But, due to the advances in information technology and the necessity to compete more effectively inside the new global markets, energy infrastructures have become increasingly automated and interlinked. The utilization of the new communication capabilities offered by information technology is a common trend for the new classes of SCADA systems. To address vulnerability, experience in secure communication systems is needed. Best practises based on securing communication system were produced by the most important research institutes in security field [1][2] [3]. Unfortunately, due to the *creativity* of potential attackers, sometimes they are not enough. The main consequence of such creativity is that the types of cyber attacks and failures are *novel* and unexpected. A more efficient mechanism is proposed in which the signatures of *normal* workings statuses of the system are recognized and deviations from normality are considered as a potential new incoming anomalies. In such a way the monitoring system will be able to produce alarms also in presence of novel and otherwise not predictable fault behaviors[4].

This work was supported by the European research IST Project SAFEGUARD

2. The cyber infrastructure

The cyber infrastructure layout chosen as test bed to test early detections of novel types of attacks/failures is the SCADA system used by the Italian Electricity Dispatching Organisation (GRTN) [5]. As shown in fig. 1, the main communication bus of such system is a Wide Area Network on which are connected, through client data concentrator devices (SIA-C), a set of Regional Control Centres.

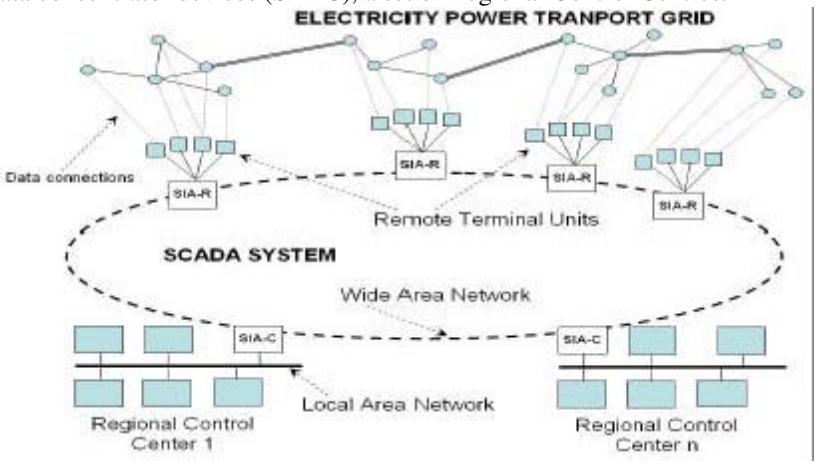


Fig 1 – Configuration of a typical SCADA system of an electricity network

Every Control Centre supervises the functioning of a part of the electricity transport grid collecting electrical data from remote concentrator devices (SIA-R) through the Remote Terminal Units. A single Remote Terminal Unit is the last digital front-end between the information system and the electrical components to be monitored. A Control Centre is composed by an operative control room where all workstations are connected together through a Local Area Network and each of them is dedicated to execute different real time processes like cyclical data acquisition, data control functions, alarm processing, event archiving and data exchanging with a National Supervisory Control Centre where day business operations are carried out.

2.1 The SCADA emulator

An experimental test bed, visualised in fig 2, was set up inside ENEA laboratories. Here an electricity "load-flow" simulator acts as source and destination of data to/from a SCADA Emulator (SE) system: its components are visualised in the figure as chequered boxes.The SCADA emulator is composed by distributed controllers that reside in different nodes of the SCADA network. An electrical simulator includes the model of a 24 buses electricity test network [6] that is proposed in the IEEE Transactions on Power Systems as an enhanced test system for use in bulk power system reliability evaluation studies.

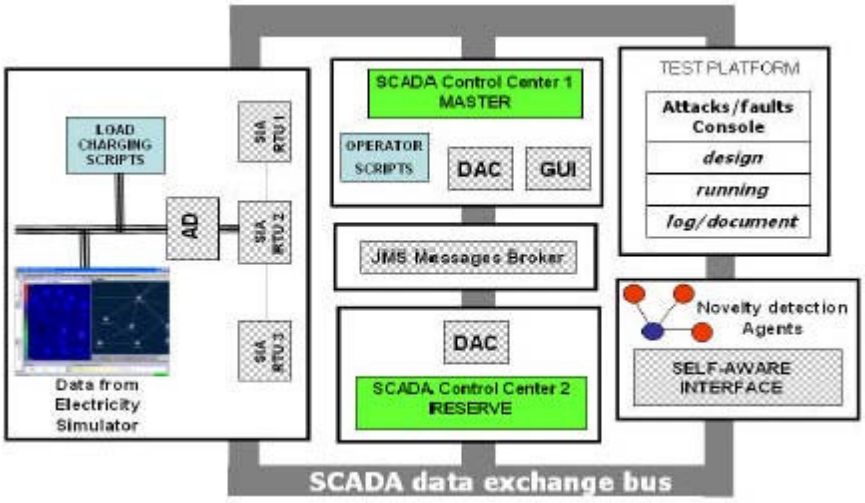


Fig. 2 – The SCADA test bed

The generated electricity data, named Tele -Measures, are voltages, active and reactive power flows, angles etc. They are acquired with a polling cycle of few seconds by the Analogue/Digital (AD) component that is the interface between the E-agorà simulator and the core of the SCADA system. Different SIA-R components may get data from one AD component. The number of SIA-Rs depends from the number of substations in the electric network. The Master Control Center contains a Graphical User Interface (GUI) and a Data Acquisition Component (DAC). Also a Reserve Control Centre is emulated; it must take the control when the Master have to be shut-down or reinitialised. From the Control Centres it is possible to send Tele-Commands, that are operator requests forexecuting manoeuvres on the network, like opening/closing breakers on the electrical lines. Each component of the test bed has been implemented in Java. They communicate each other through messages using the Java Messaging Service (JMS), so the complete system also includes a Messages Broker to route information from a sender to the right receiver. The novelty detection agents resides in a separate machine but could be instantiated to monitoring anomalies coming from the different SCADA emulator components. For this scope a special interface, named *monitoring interface* and illustrated in the next paragraph, has the duty to inform the agents about the current working status of the most important parameters.Finally a special Test Platform machine contains a special tool kit with which it is possible to design, run and log attacks and faults scenarios [12] toward the different components of the SCADA emulator.

2.2 SCADA emulator functionalities and its instrumentation interface

The most important processes carried out, during normal operations, inside the SCADA emulator are:

Request and process Tele-Measures with a polling cycle of few seconds;
Process the out of limit of a measure generating an alarm;
Process Tele-Commands generated by operator request;
Process Tele-Signals , boolean information indicating a change in the status of an electrical component like a breaker.
Process Tele-command's time-out; if, a certain time after the command request, an answering Tele-signal doesn't arrive, a time-out alarm is generated.

Every time a certain process starts inside the SCADA emulator, like a Tele-command request from operator, a corresponding *sequence of events* [13] is generated: every event of the sequence fires on the activation of different components of the SCADA emulator as illustrated in fig 3.

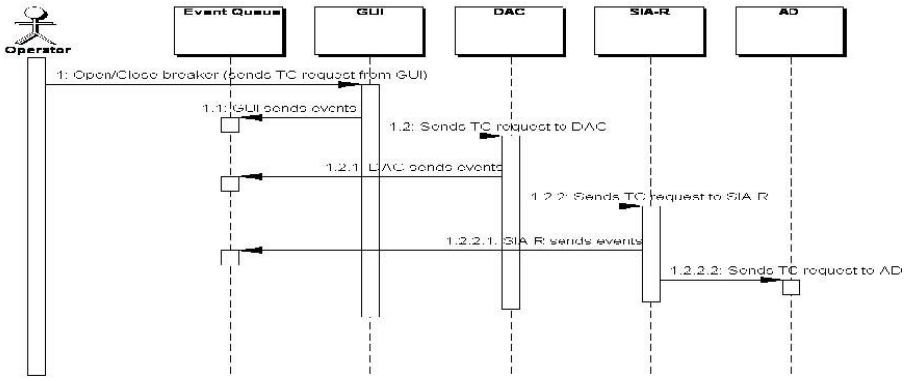


Fig. 3 – Collection of a single sequence of events

Such sequences of events, with a corresponding timing information, represents the *signature* of the executed process; the events and the associated timing information are stored inside an *events queue*. The instrumentation interface of SE contains all the events generated, during the last period, inside the SE. Events are collected as a result of a polling cycle in which, as visualised in fig 4, more processes may be executed at a same time, so that the acquisition order of the single sequence is not guaranteed. Events have anyway an associated time label.

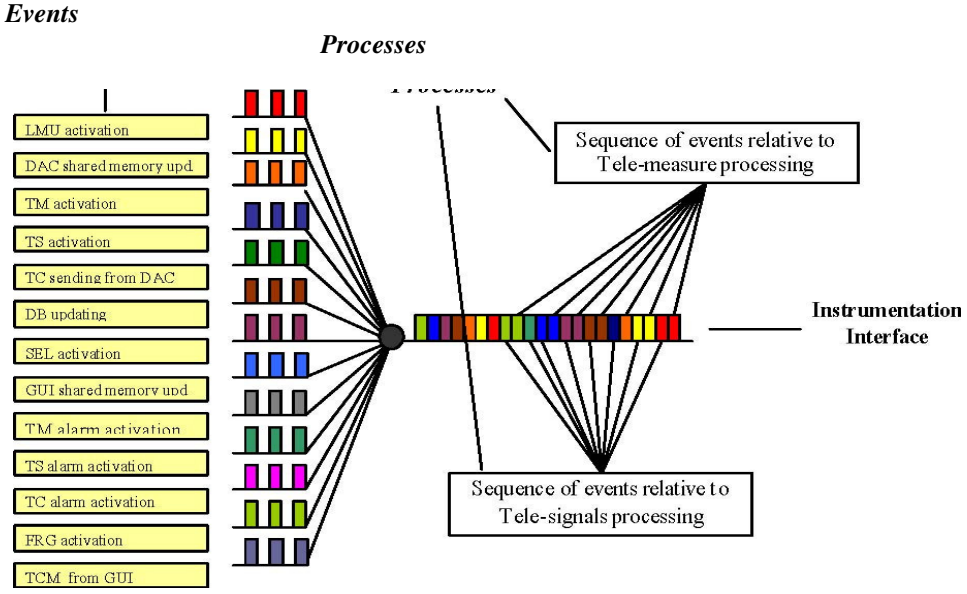


Fig 4 – Storing process of more sequences of events

During a certain sliding window the instrumentation interface contain all the information that characterize the normal behaviour of the SE during that period of time. It contains the *data behaviour* of SE from which it is possible to *learn* the characteristics of its *working model*. Making comparison between an actual behaviour with a learned one, it is possible to discover novelties in the processes carried out by the SE.

3. Novelty detection

To detect the appearance of novelties inside a working environment, knowledge about the normal working behaviour of such environment is necessary: the normal working behaviour of SE is *implicitly* contained in the data flowing inside the instrumentation interface described above. The *Self model*[7][8], functionally represents itself, the environment, and the interactions between and within each. The capacity to recognise the normal self model functioning may be defined as a *self-awareness* capacity. The capacity of a certain environment to be self-aware, is equivalent to the capacity to detect novelties emerging inside the

environment itself. The local monitoring of deviations from the normal working condition, is realized through the following *novelty detection* methods implemented inside such different components:

1 A (CBR) component for recognition anomalous *sequences of events* [13] inside the SCADA system using a *Case Base Reasoning* methodology;

2 An (NN) component for anomaly detection inside Data Sets transmitted inside SCADA system using *Auto-Encoder Neural Networks* [14][15].

3 A (DMA) component for anomaly monitoring inside Tcp/Ip exchanged data packets using a set of *Data Miner* classifiers[16].

The above three intelligent monitoring techniques are between them very different. The CBR component, that will be described more deeply in the next paragraph, is based on the availability of a certain explicit knowledge about the characteristics of the process that must be monitored. Also this one, like the others two, has the necessity to look for a certain time at the process to learn its normal behavior. The NN component is based on a special Neural network named Auto-Encoder, that has the input and output layers composed by the same number of neurons, and more additional hidden layers. This network is like a *sensor* having the input neurons connected to the input signals. If the network is well trained, when the activation values produced on the output neurons are equal to the input values it means that the normality state is recognized, otherwise an anomaly occurs. The DMA component has the primary goal to recognize anomalous TCP packets at a given port on a given host machine. The component utilize a set of classifiers generated during a learning phase; every classifier studies and makes statistics about different features of the packets headers and payloads. In the next paragraph a more detailed description of CBR component is reported.

3.1 Case Base Reasoning for novelty detection

Case-Based Reasoning (CBR) emerged from research in cognitive science from the consideration that it was not possible to build self-aware systems without a proper *management of the memory*. In CBR systems [9][10] expertise is embodied in a library of past cases, representing the system memory, rather than being encoded in classical rules. Each case is a piece of information and the case base is a format for cases representation and retrieval. Such memory makes CBR systems more self-aware than other systems, because they contain a *model of self* based on the past experience. Usually a case retrieval process selects the most similar cases to the current problem, mostly relying on nearest-neighbour techniques (a weighted sum of features in the input case is compared with the ones that identify the historical cases).

3.2 Cases representation

In our SCADA dynamic environment the Cases are lists of events representing the sequences of SCADA tasks activation for every executing process.

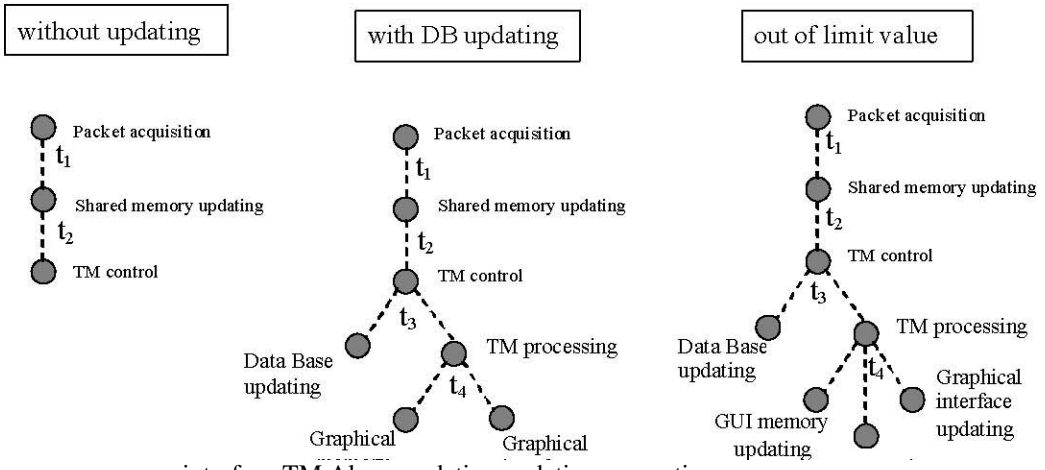


Fig 5- Modelling tasks activation process as stereotypical sequences In fig. 5 circles in the graphs represent the tasks activations (events) fired during the process, and dotted lines the temporal constraints between

them. If a node generates more than one line it means that the relative generated events could fire at the same time (asynchronously). Three different types of TM processing are visualised: the first is the default one, the second requires a DB updating process (because the TM value changed after the previous acquisition cycle), the third is similar to the second, but generates an alarm because a TM value goes out of limit. As evidenced in the figure, normal TM is a sub-sequence of TM with updating and TM with out of limit value. In fact in a SCADA system, when a cyclic measure changes, it is necessary to update the Data Base and the operator screens, and when its value goes out of limits it is also necessary to display a new alarm to the operator. These two cases are considered different because one is a sub-case of the other one also if task activation sequences are very similar. Anyway, a misclassification between them, may be a minor problem because each of them corresponds to the processing of a TM.

3.3 Similarity definition

A very important issue in every CBR system, is to define the *similarity (S)* parameter. One case (the one under study) is similar to (or it is classified by) a stereotype (the case stored in the case base) if and only if it exactly matches the sequence of events and the weakest timing constraints between the events (MlowRange, MhighRange).

$$S = \prod_{i=2}^{l} Si \quad (1)$$

In the definition (1), the sum starts from the second event because the first one has not a time delay, l is the length of the sequence and, for each event, Si is described by the function in the left part of fig 6 (event similarity function), where on x-axis is reported the time distance between the i-th event and the previous one.

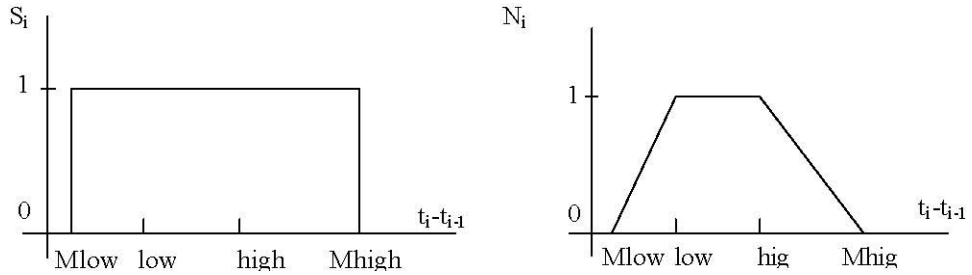


Fig 6 – Event similarity function and fuzzy normality function

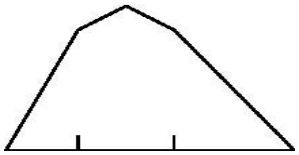
If one case can be classified by more stereotypes (as it happens with sub-cases), the highest fit is chosen, that is the case with the longest sequence. Two cases are similar if they are classified by the same stereotype. In order to check more about the normal behaviour of a case that satisfies (1) it was also introduced, as next step, the *normality* (*N*) parameter that is defined as (2) where *Ni* is defined by the following function (2) and the fuzzy function visualised in the right part of fig 6.

$$N = \frac{\sum_{i=2}^{l} Ni}{(l-1)} \quad (2)$$

The *anomaly* (*A*) parameter is then defined by:

$$A = 1 - N \quad (3)$$

A becomes 0 when all events fall in the normal range. Even if the *Ni* defined by a fuzzy-shape function is enough to detect anomaly behaviours, it was modified as described by the following in fig 7.



Ni 11-a0 ti-ti-1 Mlow low high Mhigh

Fig 7 – Fuzzy normality function with an additional sensitivity parameter "a"

This function coincides with the previous one (typical fuzzy) if $a=0$, but it allows to understand if system performances are shifting toward the border of normality, information that can be useful in some cases. Using this function, when $A<a$ all the events fall in the normal range so that "a" could be considered the first anomaly threshold. More high is the threshold (a), more detailed is the information about dispersion of measurements in the typical range. On the other side, sensitivity about anomalous values decreases while a increases.

3.4 Cases retrieval

The previous described steps (case representation, similarity and normality evaluation) are the preparatory phases of the core feature of a CBR algorithm: the *cases retrieval*. CBR, in fact, derives its power from the ability to retrieve relevant cases from its case base quickly and accurately:

- *response time* is really important in real-time systems like monitoring and control systems;
- *accuracy* is necessary because reporting utilities are useful only if they are reliable. For such anomaly detection systems the reliability is sometime jeopardized by the well known problem of the *false alarms* that will be addressed with more details in the next paragraph.

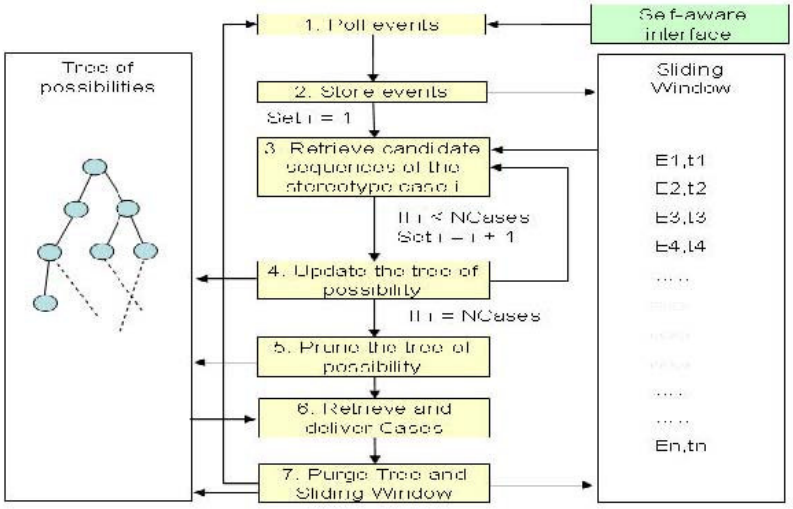


Fig 8 – Logical flow chart of CBR algorithm

The implemented CBR algorithm can be decomposed in seven logical steps as visualised in fig 8:

1 Make a cyclic polling of the events from the instrumentation interface;
2 Store the collected events in a memory buffer (temporal sliding window);
3 Scan the sliding window looking for all the sequences similar to the i-th case
4 Insert the retrieved cases inside a tree of candidate cases. Repeat the previous two steps for all the cases stored in the case base.
5 As each type of event can be involved in more cases (see fig. 4) and can satisfy timing constraints of many possibilities, a single event may be present in many candidate case, but this is physically not

acceptable. So, this step is a pruning phase aimed to recognize as valid much cases as possible by erasing (pruning) the conflicting solutions.

6 Deliver cases retrieved as valid fro m the previous step, calculating their anomaly values.

7 Purge the sliding window of the events belonging to the retrieved cases, and purge the tree of candidates cases. Return to the following polling cycle (step 1).

This algorithm was found sufficiently fast and reliable in the attack scenario implemented by using the SCADA Emulator. Anyway, the following tuning parameters can influence its performance: the *sliding window* length and the *purging time*. The sliding window is a buffer of events generated in a specified time interval. In the ideal world it should contain just one complete case. In the real one, depending on its length, it contains few complete and/or incomplete cases. The recognition of incomplete cases must continue in a next sliding window. Increasing the length of the sliding window, more cases will be present and the chance to have complete cases increases: but the system response time increases too, and over a certain level it could not satisfy the real time constraints. The purging time is the time after which an event not yet processed is purged. Increasing this time interval, gives the advantage to lost a minor number of cases, but if the case base contains only a small percentage of all possible cases a lower purging time reduces calculation and response time.

4. Spatial and timing correlation of event

In a complex infrastructure, like a distributed SCADA system, composed by an interconnected network of computerised nodes, the functionality of the whole system could not be controlled only by monitoring what is carried out in single nodes and in short time intervals. As a novel situation, detected inside a single node, cannot indicate that a general failure is in progress, and in the same way also the *absence of novelties* in such node cannot determine that no general failure is in progress. The *correlation* of events happening in a certain node of the network with events happened or that will happen in other nodes , allows to intercept phenomena not localised in a sub-set of the network. As an example we may consider the following process. The SCADA system operator sends a telecommand from the Control Centre with the objective to open a certain connection line of the electricity network. The command is firstly managed by the software and a packet is built that starts from the Control Centre and arrives to a Remote Terminal Unit (RTU) that is the most peripheral node of the distributed SCADA system, where the operator commands may arrive. The RTU manages the command and triggers the breaker opening process. To verify the absence of failures in the telecommand actuation chain it is necessary to control if, a certain time after the command has been processed by the Control Centre, it arrives and is processed by the RTU; otherwise it is possible to conclude that the command is lost. In other words it is necessary to correlate, between different locations and in subsequent times, the events that fire on the communication network. Another important difficulty for a novelty detection system is the management of *false alarm*s. In fact, some deviations from the normal working condition sometime emerge not for a real attack or failure but for temporary congestions or modifications inside the network. .

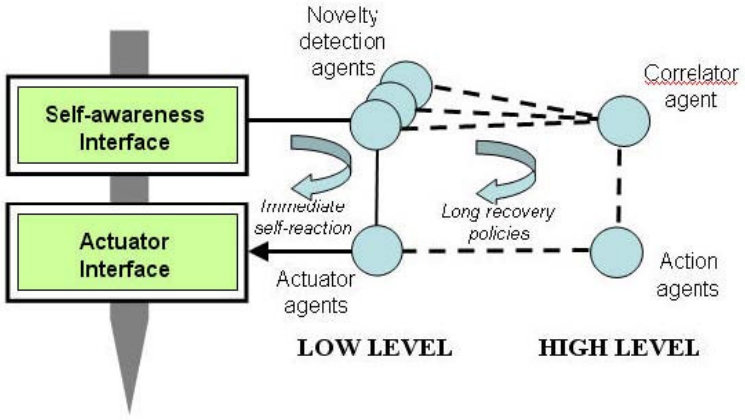


Fig 9 – The agent based architecture to integrate novelty detection algorithms

To deal with the previous described issues, the *agent based architecture* was developed as visualised in fig 9. Here the role of novelty detection agents is more evident. They make the system self-aware about *normality*; and, if a strong anomaly condition appears, they activate the *Actuator agent*, that is responsible for some immediate reactions, like the disconnection or the re-initialization of the nodes they are controlling. On the contrary, if deviation from normality is slower or the anomaly is not so evident inside a single node, a *Correlator agent* and an *Action agent*, operating at higher level, may gain same time, with the objective to study the situation during a further phase, in which many potential alarms could be verified as false. They try to correlate the events controlling if anomalies are in progress in more nodes, if they are increasing or decreasing, if they are persistent or not. The availability of such additional data, collected during a certain period of time, allows to decide what type of recovery policy is needed and if the situation must be solved at global or at local level.

5. Tests and evaluations

To test the possibility of the Correlator Agent to detect anomalies that must be confirmed by indications coming from more anomaly detectors, the test bed visualised in fig. 2 was implemented. A composite attack scenario was developed, utilising an attack/faults configuration console. The attack scenario consists in a sequence of “false tele -commands” sent from a computer machine that is connected to the SCADA emulator through the local area network. The false tele-commands are not intercepted by the operator at Control Centre but they are sent through the network toward the peripheral RTUs. Such attack produces a set of anomalous packets flowing through the communication port that may be detected by the DMA low level agent. At the same time, some worms (extraneous tasks consuming resources) are activated on the Control Centre machine emulating the way in which an intruder may conduct such anomalous hatching activities. The sequence of false tele-commands will increase packets congestion inside the network but it is not so dangerous, if in the sequence no telecommand can be “applied” on the electrical network and may produce the opening of some line electrical breakers. The detection capability of such type of attack by the low level agents can be investigated using a special system interface called “self-monitoring panel”.

5.1 Self-monitoring panel

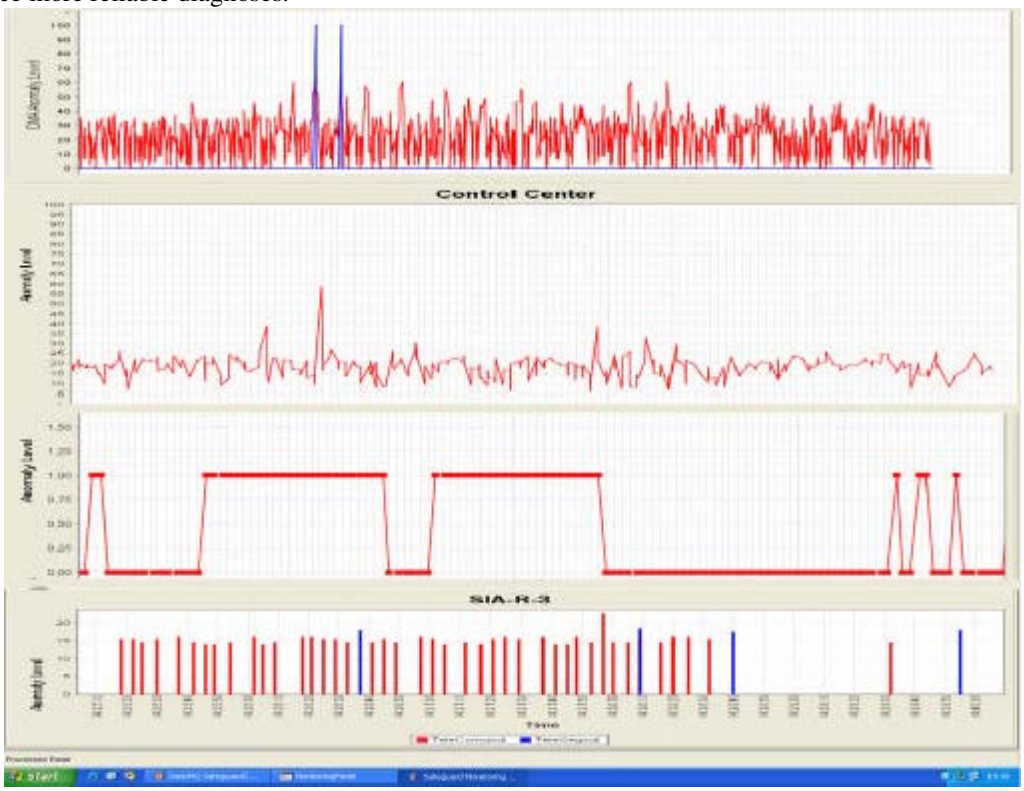
The *self-monitoring panel*, collects information about the anomaly levels detected by the novelty detection agents in different part of the network, and was designed and implemented to realise an efficient interface between the low level and the high level agents. The same interface could be also utilised to make aware a human operator about the normal working condition of the SCADA system.

In fig. 10 are visualised four snapshots of the system self-monitoring panel with anomaly data collected during the composite attack scenario described above. In the graphs of the lower part of the panel are reported anomaly values of Tele-commands and Tele-signals, intercepted at the RTU when the false tele-command sequence, for a duration of about four minutes, fired.

In the upper first graph is shown the anomaly graph detected by DMA low level agent, where is possible to evidence an increasing of 15/20% in the anomaly values of many detected packets. In the upper second graph is reported anomalies detected in the processing of Tele-measures at Control Centre (50% more respect to the normal in some cases) caused mainly by anomalous resources consuming inside the machine. The tele -measure graph shows a constant average anomaly level of 20% and a maximum fluctuations of values of about 10%. This stable condition indicates a normal working condition of the system during tele-measures processing. Deviations from the normal anomaly level or from the normal maximum fluctuation values are easily and early intercepted and indicate some incoming anomalous conditions.

In the upper third graph is evidenced how the NN agent detect the anomalous conditions in the data sets acquired by the RTU when a tele-command was “applied” and consequently the network status changed. Some time before the end for the attack period the operator re-close the opened breaker and the anomaly status was resettled.

Generally an anomaly state detected by a single low level agent can be a starting point to initiate a diagnostic reasoning, but if no other “correlated” anomaly is found coming from another agent is an indication of the possibility of a false alarm. Correlation is for this reason a very important mechanism to produce more reliable diagnoses.



6. Conclusions and further developments

On the basis of the executed tests, novelty detection agents, that make use of an event based self-model, seem to be a promising technique for early discovering of incoming malfunctioning that initiate at a certain part of the network, and that could not be foreseen with different approaches. It seems anyway very important, to increase the efficiency of such systems, providing them with higher level agents able to exploit the results of the low level agents for detecting and confirming failures at more general level. To address this issue correlation mechanisms will be developed for a correlation agent able to implement space and time based diagnostic processes. In addition, to avoid a too high number of false alarms, it seems also

important to apply the recovery policies by evaluating, during a certain time window, the trend and the persistence of the detected anomalies. At the present the functionalities of these high level agents are under development and will be tested in the same SCADA emulator environment. In the same way, as the instrumente interface is used by the low level agents to understand and detect anomalies in single parts of the physical network, the self-monitoring panel will be used by the high level agents to detect anomalies in the behaviour of many low level agents controlling different part of the physical network.

References

[1] P. Oman, E. O. Schweitzer III, and J. Roberts, Safeguarding IEDs, Substations, and SCADA systems against Electronic intrusions, Schweitzer Engineering Laboratories Inc., Pullman WA, USA, Technical report, April 2001.

[2] R. J. Ellison, D. A. Fisher, R. C. Linger, H. F. Lipson, T. A. Longstaff, and N. R. Mead "Survivability: protecting your critical systems", IEEE Internet Computing, Vol. 3, Issue 6, Nov-Dec. 1999. Pp. 55 – 63.

[3] Schneier, B., Secrets and Lies: Digital Security in a Networked World, John Wiley & Sons, August 2000.

[4] Somayaji, A.B. Operating System Stability and Security through Process Homeostasis. Ph.D. thesis. University of New Mexico, Alberquerque, NM, June 2002. Available from http://www.cs.unm.edu/ soma/pH/.

[5] A. Serrani, M. Mocenigo, M. Sforna, The Dispatching Organisation of the Italian Independent System Operator, HK CIGRE, 4th Symposium on Power System Management, Cavtat-Croazia, 22-25 Oct. 2000.

[6] C. Grigg, P. Wong et al., The IEEE Reliability Test System – 1996, IEEE Transaction on Power Systems, Vol. 14, No. 3, p.p 1010- 1020, August 1999

[7] T. Metzinger, *The Self-Model Theory of Subjectivity* MIT Press, (2003), pp. 699, ISBN: 0-262-13417-9

[8] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for UNIX processes. In *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, *pp 120-128*, Los Alamitos, CA, 1996. IEEE Computer Society Press.

[9] C. Balducelli, F. Brusoni: "A CBR Tool to Simulate Diagnostic Case Based Operator's Model", Proceedings of 8th European Simulation Symposium, p.p. 298-301, Oct. 1996, Genoa, Italy.

[10] Sheng Li, Quiang Yang (2001), An Agent System that Integrate Case-Base Reasoning and Active Databases, *International Journal of Knowledge and Information System* (2001) 3:225:251, Spriger-Verlag

[11] Wenke Lee, Salvatore Stolfo, and Patrick Chan. Learning patterns from unix process execution traces for intrusion detection. In Proceedings of the AAAI97 workshop on AI methods in Fraud and risk management, 1997.

[12] Balducelli C.(2003), Modelling Attack Scenarios against Software Intensive Critical Infrastructures, *10th Annual Conference of The International Emergency Management Society*, Sophia-Antipolis, Provence, France, June 3-6, 2003.

[13] Terran L., Brodley C. E.(1999), Temporal sequence learning and data reduction for anomaly detection, *ACM Transactions on Information and System Security (TISSEC)*, v.2 n.3, p.295-331, Aug. 1999

[14] Timusk M.A., Meckefske C.K. (2002), Applying Neural Network based Novelty Detection to Industrial Machinery, *Proceedings of KES2002 Knowledge-Based Intelligent Information Engineering Systems & Allied Technologies, IOS Press Ohmsha*, Podere d'Ombriano Sep. 2002

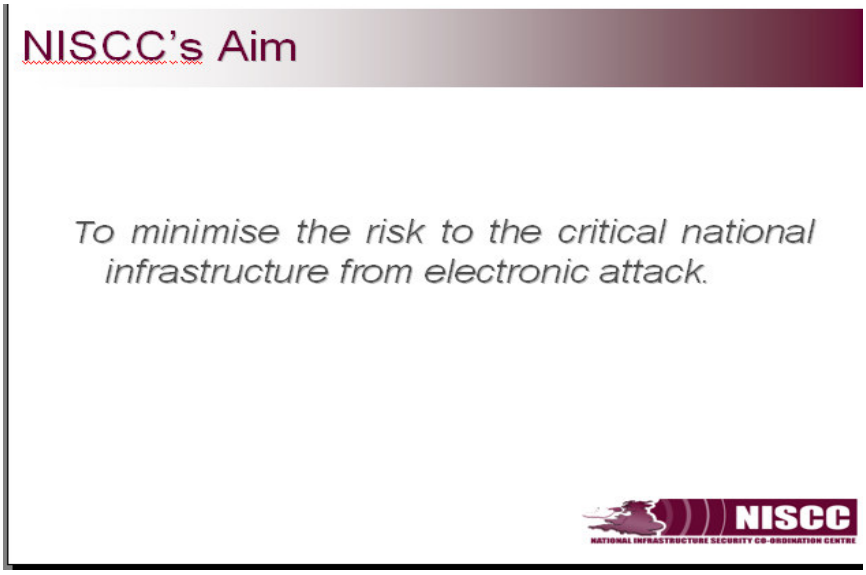
[15] M. Martinelli, E. Tronci, G. Dipoppa and C. Balducelli, Electric Power System Anomaly Detection Using Neural Networks, *Proceedings of KES2004 Knowledge-Based Intelligent Information Engineering System* , New Zeland, 2004

[16] Witten H., Frank E., Data Mining, *Morgan Kaufmann Publisher*, 2000, S. Francisco, C.A., USA

Critical Information Infrastructure Protection: The Role of the UK National Infrastructure Security Coordination Center

Mike Corcoran,

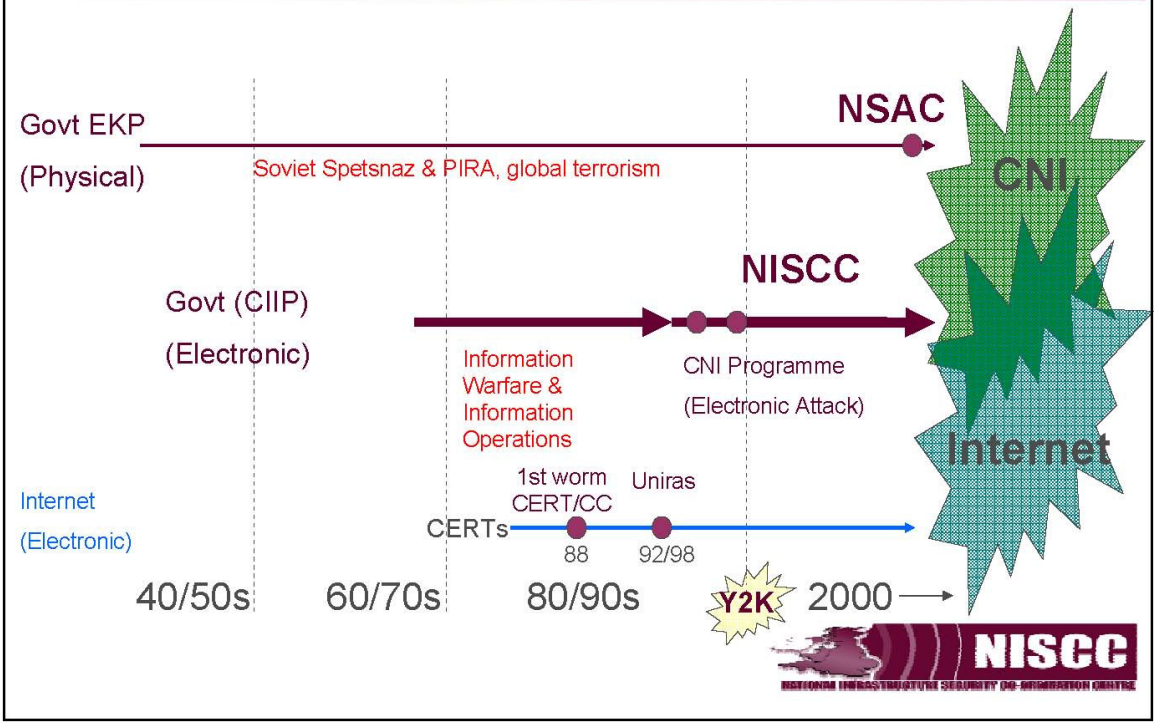
UK National Infrastructure Security Co-ordination Center,
mikec@niscc.gov.uk



Changing times…

NISCC

Protective Security Time Line



The CNI Sectors

- Telecommunications
- Energy
- Finance
- Government & Public Services
- Water and Sewerage
- Health Services
- Emergency Services
- Transport
- Hazards
- Food

NISCC
NATIONAL INFRASTRUCTURE SECURITY CO-ORDINATION CENTRE

An Interdepartmental Centre

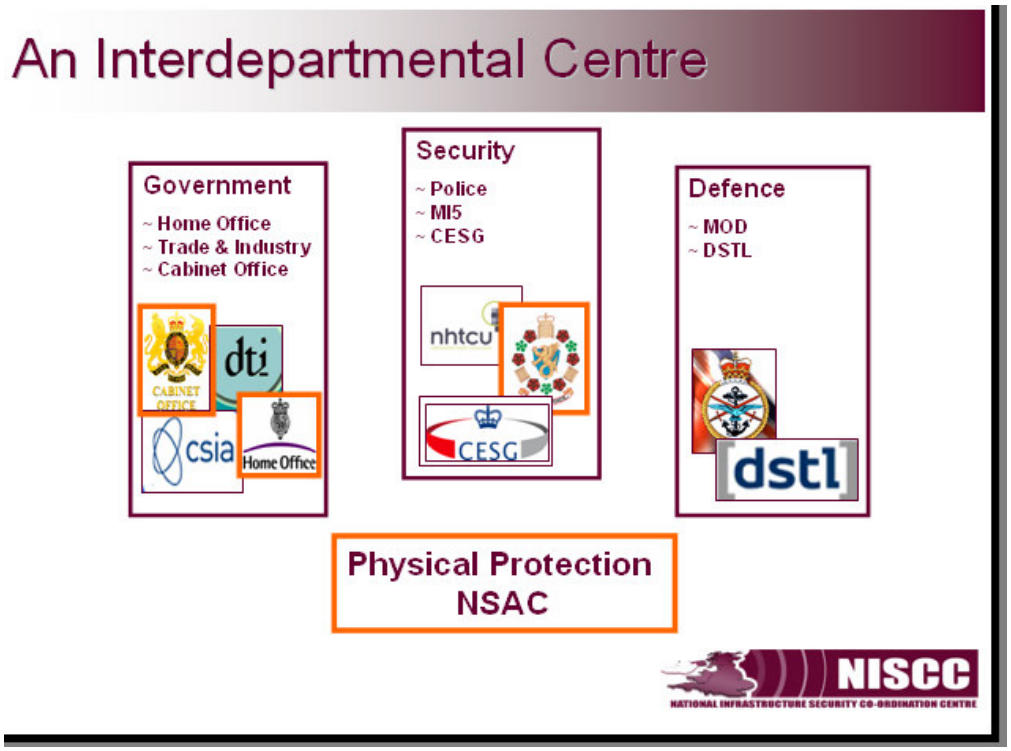
Government
~ Home Office
~ Trade & Industry
~ Cabinet Office

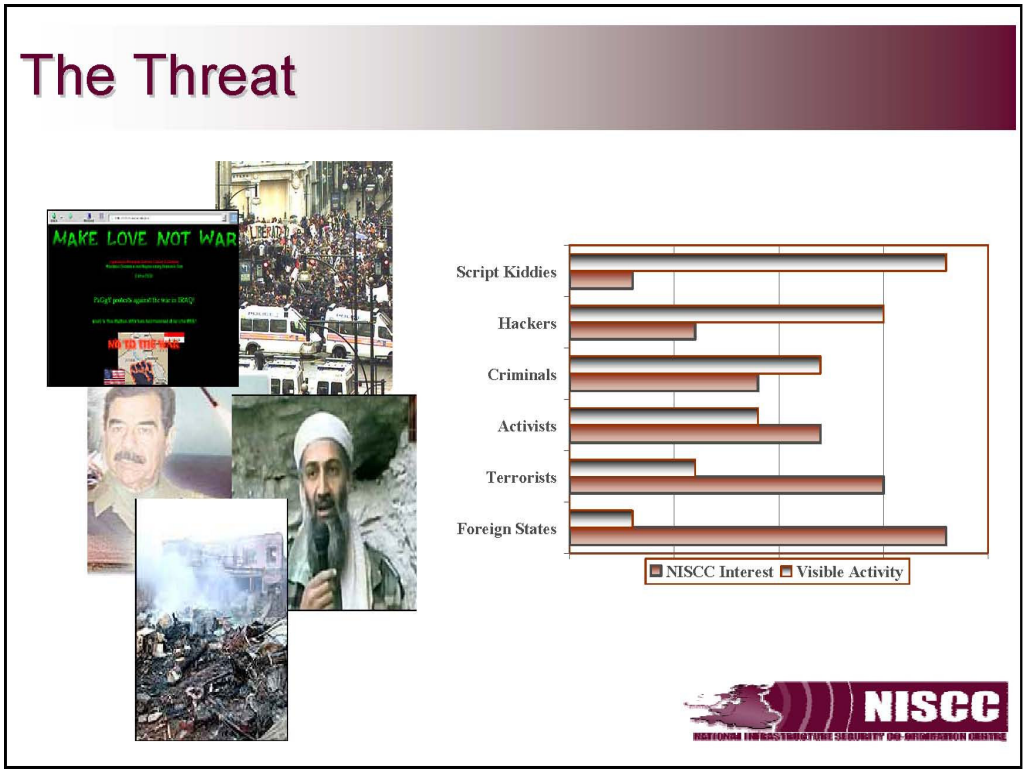
Security
~ Police
~ MI5
~ CESG

Defence
~ MOD
~ DSTL

contribute to

NISCC


An Interdepartmental Centre
Government
~ Home Office
~ Trade & Industry
~ Cabinet Office
CABINET OFFICE
dti
csia
Home Office
Security
~ Police
~ MI5
~ CESG
nhtcu
CESG
Defence
~ MOD
~ DSTL
dstl
Physical Protection
NSAC
NISCC
NATIONAL INFRASTRUCTURE SECURITY CO-ORDINATION CENTRE

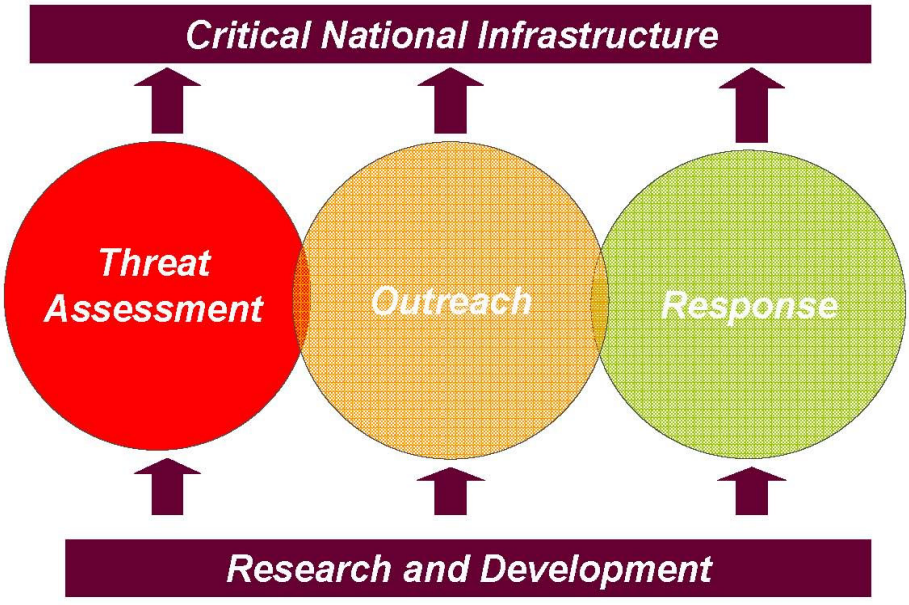

The Threat
MAKE LOVE NOT WAR
Script Kiddies
Hackers
Criminals
Activists
Terrorists
Foreign States
NISCC Interest
Visible Activity
NISCC

However…

- We do not have enough information
- The threat level can change dramatically
- The insider threat

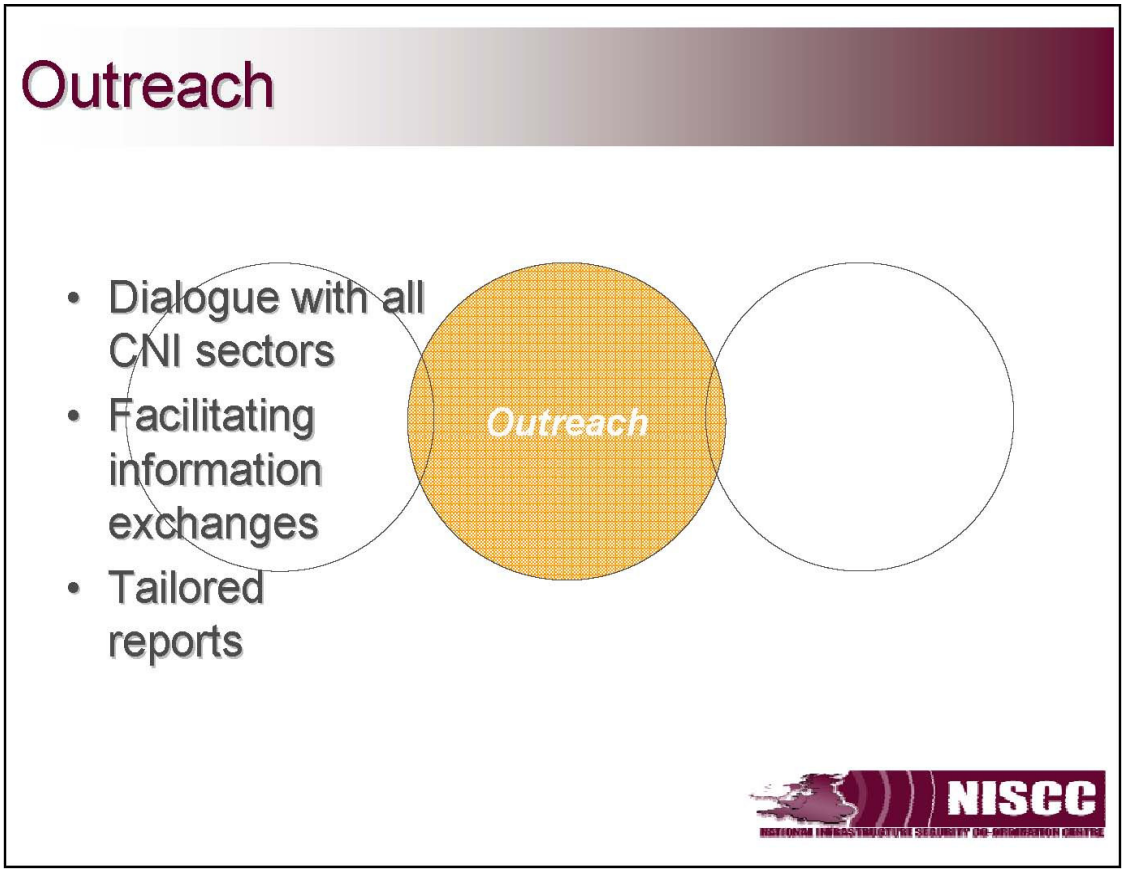
NISCC
NATIONAL INFRASTRUCTURE SECURITY CO-ORDINATION CENTRE

How NISCC works



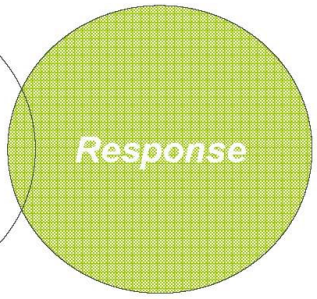
NISCC


Investigating and Assessing the Threat
Threat Assessment
• Making best use of technical, human and open sources to investigate
• Analysis and assessment
• Reports and specific threat assessments
• Disruptions
NISCC
NATIONAL INFRASTRUCTURE SECURITY CO-ORDINATION CENTRE


Outreach
• Dialogue with all CNI sectors
• Facilitating information exchanges
• Tailored reports
Outreach
NISCC

Response

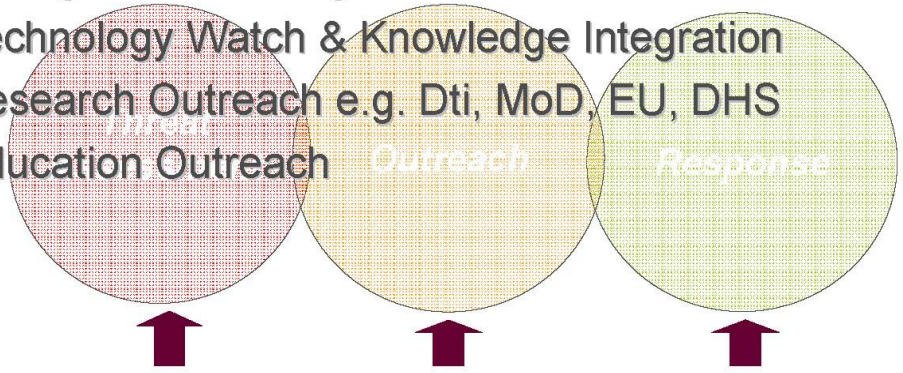
- Briefings and alerts via UNIRAS
- Responsible disclosure of vulnerabilities
- Assistance with recovery from direct attacks



NISCC

e-Science R & D

- Key Technologies e.g. MPLS, VoIP, SS7, SCADA, ATN.
- Emergent Technologies e.g. Genome, RF/DEW, GPS.
- Technology Watch & Knowledge Integration
- Research Outreach e.g. Dti, MoD, EU, DHS
- Education Outreach

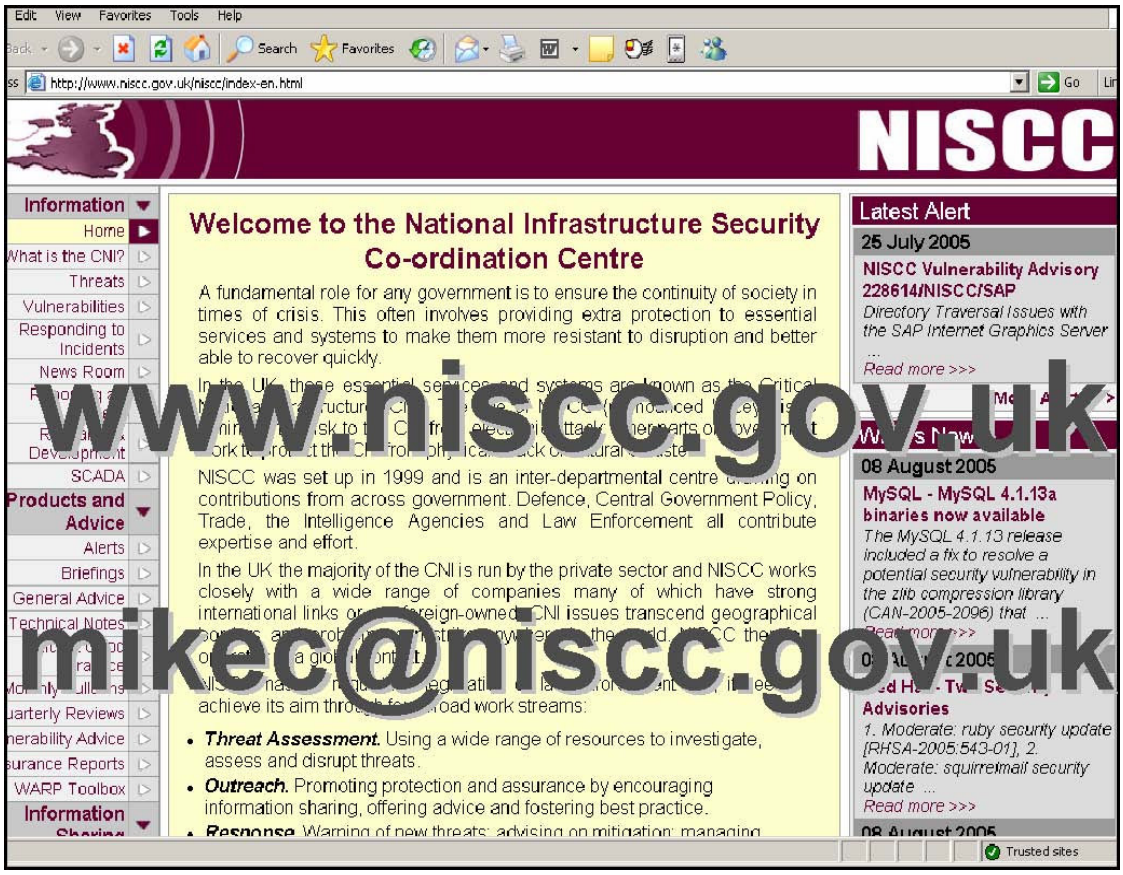


NISCC

Beyond the CNI…

- NISCC alerts and warnings already go wider than the CNI.
- NISCC vulnerability disclosure process now world class. Many companies ask us to handle this for them.
- Concept of WARPs (Warning Advice and Reporting Points) for non CNI organisations.
- ITSafe programme, for general public

NISCC
NATIONAL INFRASTRUCTURE SECURITY CO-ORDINATION CENTRE



Methodology for Identifying Near-Optimal Interdiction Strategies for a Power Transmission System

Vicki M. Bier (1), Eli R. Gratz (2), Naraphorn J. Haphuriwat (2), and Wairimu Magua (2), Kevin R. Wierzbicki (3)

(1) Department of Industrial and Systems Engineering, University of Wisconsin-Madison
Madison, WI 53711, U.S.A.
http://www.engr.wisc.edu/ie/faculty/bier_vicki.html

(2) Department of Industrial and Systems Engineering, University of Wisconsin-Madison
Madison, WI 53711, U.S.A.

(3) Department of Electrical and Computer Engineering, University of Wisconsin-Madison
Madison, WI 53711, U.S.A.

Abstract

Previous methods for assessing the vulnerability of complex systems to intentional attacks or interdiction have either not been adequate to deal with systems in which flow readjusts dynamically (such as electricity transmission systems), or have been complex and computationally difficult. We propose a relatively simple, inexpensive, and practical method ("Max Line") for identifying promising interdiction strategies in such systems. The method is based on a greedy algorithm in which, at each iteration, the transmission line with the highest load is interdicted. We apply this method to sample electrical transmission systems from the Reliability Test System developed by the Institute of Electrical and Electronics Engineers, and compare our method and results with those of other proposed approaches for vulnerability assessment. We also study the effectiveness of protecting those transmission lines identified as promising candidates for interdiction. These comparisons shed light on the relative merits of the various vulnerability assessment methods, as well as providing insights that can help to guide the allocation of scarce resources for defensive investment.

1. Overview

Electric power transmission grids are an important component of the modern economy (Electricity Consumers Resource Council, 2004). We rely on electricity for communications, light, water, transportation, heating, and industry, among other critical uses of power. As a result, numerous researchers have studied the risk of electric blackouts; see for example Carreras et al. (2002), Chen et al. (2001), Liao et al. (2004), Mili et al. (2004), and Phadke (2004). Vulnerability studies have been recognized as being important in assessing the reliability of critical infrastructure and helping to guide defensive investments since even before the terrorist attacks on September 11th, 2001 (North American Electric Reliability Council, 2001); see for example Guzie (2000) for an application of vulnerability analysis to military systems, and Ezell et al. (2000a, 2000b, 2001) for applications to water systems. Methods for assessing and improving the vulnerabilities of critical infrastructure have also been the focus of substantial government research programs; see for example Los Alamos National Laboratory (2004).

One of the most promising approaches for vulnerability assessment is that proposed by Apostolakis and Lemon (2005), since it explicitly takes into account the complex networked structures of many infrastructure systems. However, that approach is limited to distribution systems (with one-directional flows), in which the consequences of interdicting a given line can be determined in a straightforward manner. It is important to extend this methodology to transmission systems, since Zimmerman et al. (2005) state that the majority of electricity outages and terrorist attacks on electricity systems involve damage to

transmission equipment. This will require some method of accounting for the fact that transmission systems can have bi-directional flows, and that flows can therefore be reconfigured dynamically after one or more transmission lines have been removed.

Salmeron et al. (2004) model interdiction of electricity transmission system using a non-linear program. However, their formulation of the problem is difficult to solve, since it involves a nested optimization (minimization of costs to determine power flows on the network, with maximization of damage to identify an interdiction strategy), with the outer loop entailing maximization of a convex rather than a concave function. They are able to solve their model only using a heuristic algorithm, so the resulting interdiction strategies are not known to be optimal. The non-linear programming approach also seems impractical for use on large problems.

In extending the work of Apostolakis and Lemon to transmission systems, we initially considered the option of taking out transmission lines randomly, in an approach similar to that applied by Schaefer and Bajpai (2004, 2005; see also Bajpai and Schafer, 2003) in the context of load-bearing members of buildings or other structures. However, while potentially useful in anticipating "unforeseen hazards" in general, that approach did not seem adequate for modeling the effects of terrorist actions or other intentional malevolent acts, where presumably some intelligence is devoted to determining which elements to attack. It also had the potential to be computationally costly, if large numbers of random "attacks" were needed to identify a few that were seriously damaging. Therefore, we decided to take out transmission lines in decreasing order of load. Albert et al. (2004) have indicated that "connectivity loss is significantly higher" when interdiction of transmission-system components is in decreasing order of load rather than random.

This study offers a viable method of identifying strategies that result in substantial unmet demand for electricity. Our method extends the work of Apostolakis and Lemon (2005) from distribution networks to transmission networks, yielding results that compare favorably to those of Salmeron et al. (2004). The methodology reflects the dynamic nature of transmission grid power flow, but is simple enough to implement in practice even for relatively complex systems. We use the same nested optimization approach as Salmeron et al., but our method avoids their computational difficulties, since the outer maximization loop is trivial and can be solved by inspection.

2. Case Study and Approach

We apply our method to the IEEE Reliability Test System – 1996 (RTS-96; Reliability Test System Task Force of the Application of Probability Methods Subcommittee, 1999), which is designed to be representative of typical transmission systems. We analyze both the IEEE One Area RTS-96 and the IEEE Two Area RTS-96 (which combines two separate areas using three interconnections). We model the IEEE One Area RTS-96 as a network consisting of 24 nodes and 38 arcs, and the Two Area RTS-96 as a network consisting of 48 nodes and 79 arcs. We base our analysis on decoupled load (DC) flow with optimal dispatch.

Our approach is based on three nested algorithms: a load-flow algorithm; a Max Line interdiction algorithm; and a hardening algorithm. The load-flow algorithm is used to determine optimal DC power flow dispatch on the transmission network, both before and after any interdiction of transmission lines. The Max Line interdiction algorithm identifies the transmission line transporting the most DC flow (to be removed from the network by supposed malevolent attackers), after which flows are re-optimized using the load-flow algorithm. We refer to each cycle of interdiction and re-optimization as an iteration. The hardening algorithm then simulates a system upgrade by hardening (making invulnerable) some of the transmission lines identified for interdiction by the Max Line algorithm. After hardening has been implemented, the Max Line algorithm can then be applied in successive iterations to identify "next best" interdiction strategies. These algorithms are described in Sections 3-5, respectively.

For simplicity, we consider only the interdiction of electric transmission lines (arcs), not nodes (such as transformers). We compare our methods and results to those of Salmeron et al. (2004) and Apostolakis and Lemon (2005).

We now introduce the following notation used in describing our algorithms:

B	set of nodes in the network, indexed by i
L	set of lines in the network, indexed by k
G_i	generation at node i
L_i	load supply at node i
$L_{i, demand}$	load demand at node i
$L_i(t)$	load supply at node i after iteration t of the Max Line algorithm
F_k	negative or positive power flow on line k (to reflect bi-directional flow)
$F_{k, max}$	maximum power flow permitted on line k (in absolute value)
F	vector of F_k for all k ϵ ***L***
P_i	total power at node i (given by $G_i - L_i$)
P	vector of P_i for all i ϵ ***B***
$W_{gen, i}$	cost of generation at node i
$W_{shed, i}$	cost of load shedding at node i
M	DC load flow matrix relating line flows F to power levels P
k*(t)	index of the line with the highest absolute value of power flow at iteration t of the Max Line algorithm
K(t)	set of lines attacked in iteration t of the Max Line algorithm
A	ordered set of (sets of) attacked lines ***K(t)***
A(s)	ordered set of (sets of) attacked lines after iteration s of the hardening algorithm
H	set of hardened lines

3. Load-Flow Algorithm

To simulate power flows on the network, we use a DC load-flow model (Salmeron et al., 2004; Carreras et al., 2002). This optimization problem minimizes the cost function

$$\sum (G_i W_{gen, i} - L_i W_{shed, i}) \quad (1)$$

subject to the following constraints:

$$0 \leq G_i \leq G_{i, max} \quad (2)$$

$$-L_{i, demand} \leq -L_i \leq 0 \quad (3)$$

$$-F_{k, max} \leq F_k \leq F_{k, max} \quad (4)$$

$$F = MP \quad (5)$$

For any given set of available lines, both generation and load flows are assumed to be determined as the solution to the above optimal dispatch problem. The objective is to minimize the combined cost of generation and unmet demands. Constraint (2) ensures that no generator exceeds its maximum power output. Constraint (3) ensures that the load supplied at any given node does not exceed the corresponding demand. Constraint (4) ensures that power flows on the lines remain within safe margins. Constraint (5) is a matrix equation relating the vector of power levels at each node with the vector of power flows on each line through the constraint matrix M. For details, consult Carreras et al. (2002) or Salmeron et al. (2004).

In general, the costs or weights, $W_{gen, i}$ and $W_{shed, i}$, can take on different values at each node, representing different prices at each generator and different levels of importance of each load respectively. However, in our case, we set each generator price to 1 and each load importance to 100, as in Carreras et al. (2002).

4. The Max Line Interdiction Algorithm

We assume that the attacker uses a greedy algorithm where, at each iteration, the line with the maximum flow is effectively disabled or removed from the system. The load-flow algorithm is then run to compute the optimal power dispatch on the revised system. The interdiction algorithm is terminated after a predetermined number of steps. The algorithm can be summarized as follows:

Step 1: The system is initialized at iteration t = 0, at which time the set ***A*** is empty. The set ***H*** is also empty, unless the hardening algorithm has already been run one or more times, in which case ***H*** contains the lines selected for hardening as a result of that algorithm.

Step 2: The load-flow algorithm is run, and optimal dispatch is determined. The resulting load shed or unmet demand (which may be zero), $L_{i,\,demand} - L_i(t)$, at each bus $i \in \boldsymbol{B}$ is recorded. The set ***K(t)*** is also initialized to be empty.

Step 3: The line k*(t) for which the absolute value of power flow is given by $\{\max |F_k(t)|: k \in \boldsymbol{L\text{-}H}\}$ is found, and k*(t) is added to ***K(t)***. If there is more than one such line, k*(t) is chosen at random from those lines whose absolute value of power flow is equal to $\{\max |F_k(t)|: k \in \boldsymbol{L\text{-}H}\}$. Any lines in close geographical proximity to k*(t) are also added to ***K(t)***.

Step 4: The lines in ***K(t)*** are removed from the network by setting $F_{k,\,max}$ to zero for all $k \in \boldsymbol{K(t)}$. These changes remain in effect through all subsequent iterations of the interdiction algorithm. The set ***K(t)*** is also added as the t^{th} element of the ordered set ***A***.

Step 5: The index t is incremented by 1, and the algorithm returns to Step 2, unless it has reached the pre-determined maximum number of iterations.

5. Hardening Algorithm

The hardening algorithm can be run after the Max Line interdiction algorithm to simulate an "improvement" of the system to reduce the consequences of an attack. In this case, the interdiction algorithm is rerun after each successive run of the hardening algorithm to investigate the effectiveness of the postulated system hardening.

The hardening algorithm is summarized below:

Step H-1: The system is initialized at iteration s = 0, with the set ***H*** empty.

Step H-2: The Max Line interdiction algorithm is run for some number of iterations t, resulting in an ordered set ***A(s)*** consisting of t sets of attacked lines.

Step H-3: The first n elements of ***A(s)***, ***K(1)*** through ***K(n)***, are chosen for hardening, and added to the set of hardened lines ***H***. (In the application of this algorithm in section 6, we choose n=5 for the one-area network and n=10 for the two-area network.) The hardened lines are no longer candidates for interdiction, as shown in Step 3 of the Max Line interdiction algorithm.

Step H-4: The hardening index s is incremented by 1, and the program returns to step H-2, unless it has reached the maximum number of hardening iterations.

6. Results

In Figure 1, we graph the load shed pattern that would result from the first fourteen iterations of the Max Line algorithm applied to the one-area system. Each of the iterations on the horizontal axis represents the removal of a line, or two or more lines in close geographical proximity (as described in RTS-96), from the network. The corresponding value on the vertical axis shows the unmet load after optimal re-dispatch of power flow on the remaining lines.

In our proposed interdiction plan, the first three iterations of the algorithm (leading to the interdiction of four transmission lines) in the one-area system result in a 44% loss of load, indicating that attacking only 11% of the transmission lines in the system would result in significant unmet demand. The first nine iterations (corresponding to 11 transmission lines, roughly a third of the lines in the system) result in a 56% loss of load. Removing additional lines does not result in substantial additional loss of load, because the system is already largely unconnected and serving primarily local loads by this point.

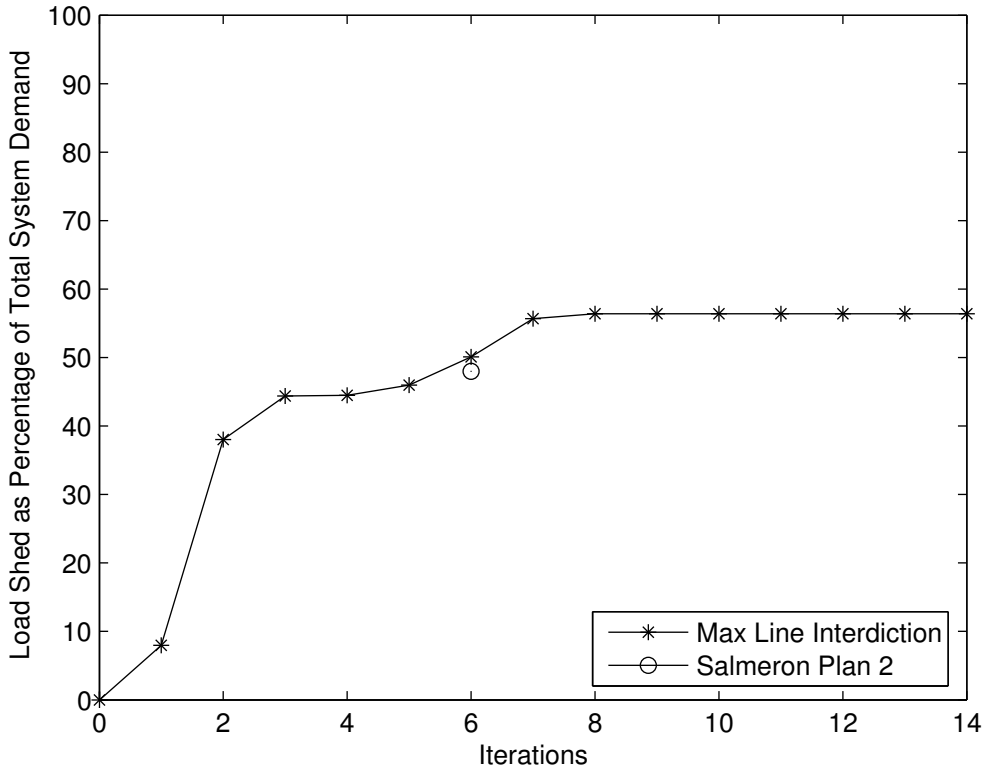


Figure 1: Load shed comparison between the Max Line interdiction strategy and Plan 2 of Salemeron et al. for the One Area RTS-96.

We now compare the results of our methodology with those obtained by Salmeron et al. (2004), who developed two candidate interdiction plans for the IEEE One Area RTS-96. Since we do not consider the interdiction of substations in our method, we therefore compare our results only to the line interdiction strategy (Plan 2) developed by Salmeron et al. Nine lines are interdicted in Plan 2 (corresponding to six sets of lines in close geographical proximity).

As illustrated in Figure 1, Plan 2 of Salmeron et al. (2004) results in shedding about 48% of the total system demand. By contrast, the Max Line algorithm results in a 50% load shed after six iterations (corresponding to eight lines). Note, by the way, that the transmission lines interdicted in the strategy proposed by Salmeron et al. differ from those interdicted in our strategy.

We also study the IEEE Two Area RTS-96. Plan 3 proposed by Salmeron et al. sheds approximately 44% of system load after the removal 11 sets of lines in close geographical proximity (corresponding to 17 transmission lines). By contrast, the Max Line algorithm results in 45% load shed after eleven iterations (corresponding to fifteen lines).

Thus, the Max Line interdiction strategy reasonably approximates the load shed by the near-optimal attack plan developed by Salmeron et al. (2004). Note that Salmeron et al. do not weight all transmission-system components equally. Therefore, it is possible that their algorithm would perform better than ours if both algorithms were applied using the same weights. However, Salmeron et al. specifically state that their weights are chosen to improve the efficiency of their algorithm. In any case, we find the performance of the two approaches to be remarkably close.

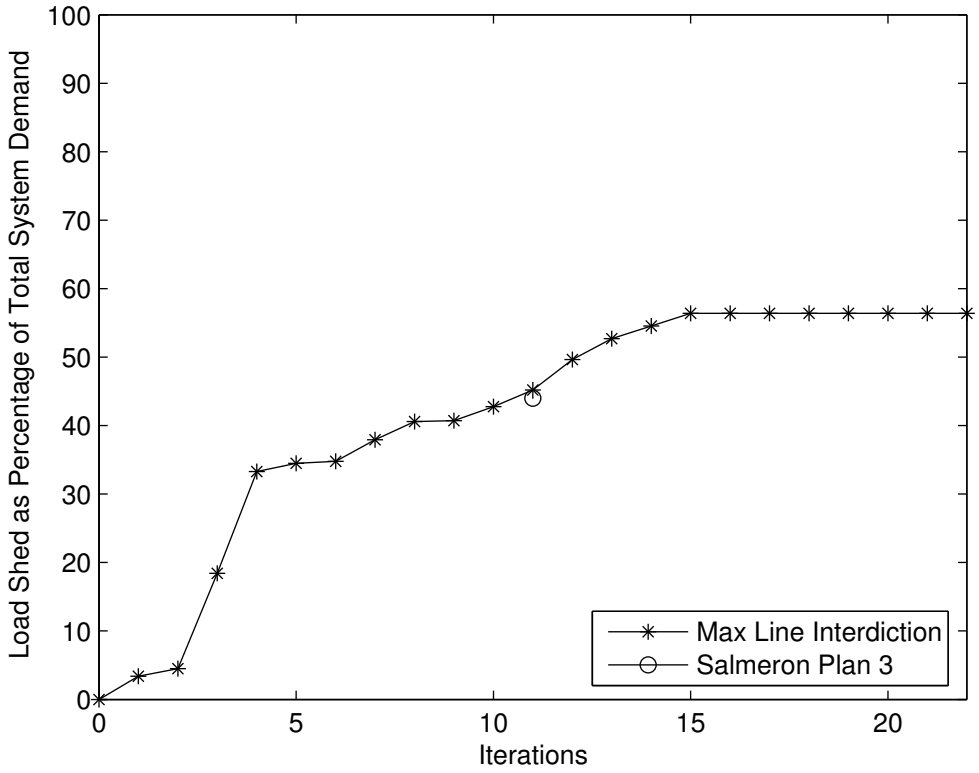


Figure 2: Load shed comparison between the Max Line interdiction strategy and Plan 3 of Salmeron et al. for the Two Area RTS-96.

We now compare the Max Line strategy against random removal of lines from the one-area transmission system. In this example, the first five random iterations (corresponding to seven transmission lines) shed only 9% of the total system demand. By contrast, the first five iterations of the Max Line algorithm (corresponding to seven transmission lines) result in a loss of approximately 46% of the total system demand, as shown in Figure 3. We conclude that random interdiction appears to be an inefficient strategy for identifying vulnerabilities (although even random interdiction can have a significant effect on system connectivity if a sufficiently large number of lines are interdicted, as shown in Figure 3).

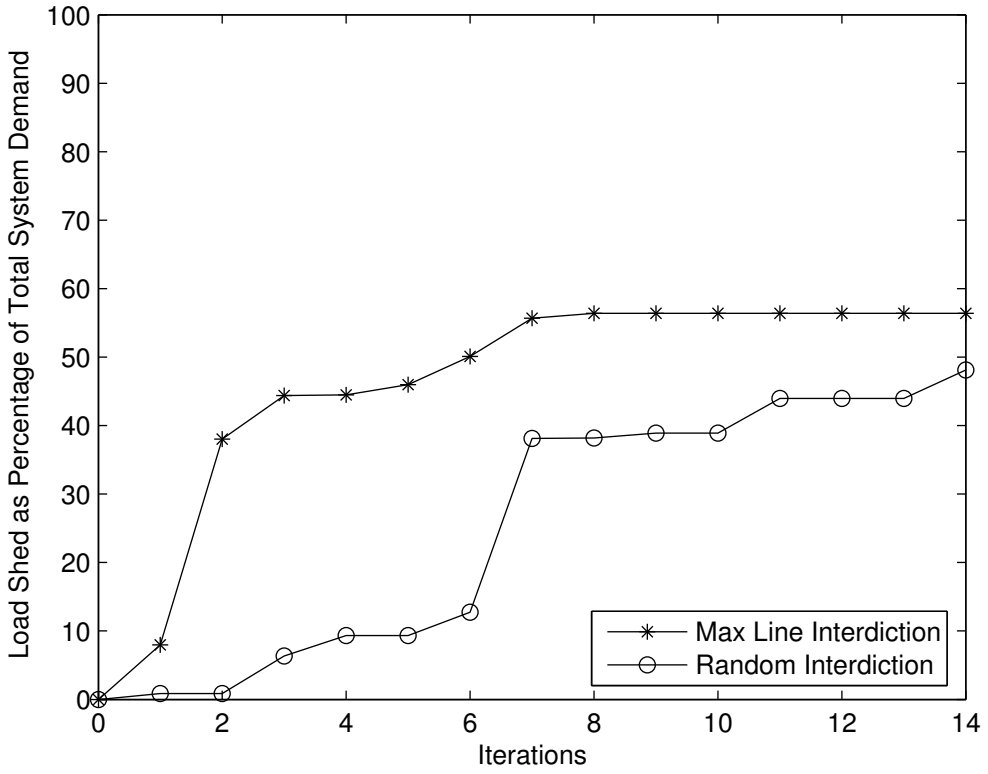


Figure 3: Load shed comparison between the Max Line interdiction strategy and random removal of transmission lines for the One Area RTS-96.

Next, we apply the hardening algorithm to simulate an upgrade of the system, as described in Section 5. This examines the impact of protecting attractive targets in both the IEEE One Area RTS-96 and the IEEE Two Area RTS-96. H0 represents the original interdiction strategy, as shown in Figure 4 or Figure 5, as appropriate. Strategies H1, H2, and H3 show the interdiction strategies obtained after each of three iterations of the hardening algorithm.

For the IEEE One Area RTS-96, strategy H0 (with no hardening) results in a loss of 56% of the total system demand. By contrast, strategy H3, after hardening 15 sets of transmission lines in close geographical proximity (approximately 39% of all lines in the system), still results in a loss of 42% of the total system demand.

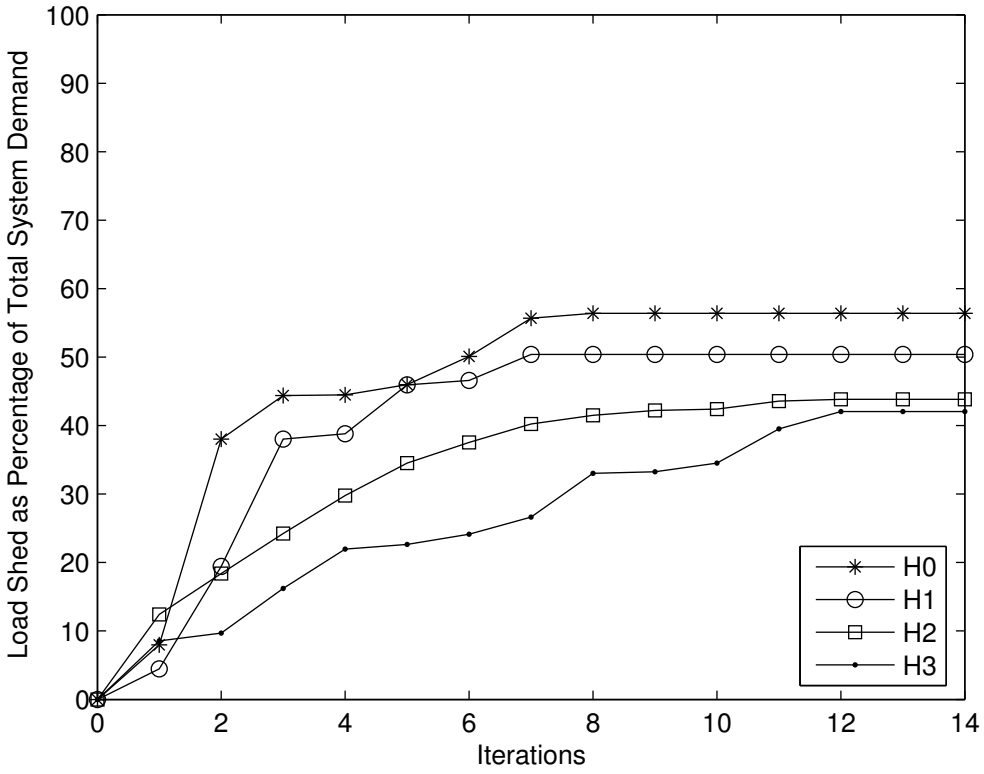


Figure 4: Interdiction strategies generated after hardening of the One Area RTS-96.

We now study the same cycle of hardening and interdiction for the IEEE Two Area RTS-96. The results are shown in Figure 5. Strategy H0 results in a loss of 56% of total system demand. Strategy H3, after hardening 39% of the transmission lines in the system, results in a loss of 39% of total system demand.

Our results cast doubt on the observation by Salmeron et al. that "By considering the largest possible disruptions, our proposed plan will be appropriately conservative." In fact, we observe that hardening even a significant percentage of the transmission lines in the system does not dramatically diminish the load that can be shed as the result of an intelligent attack. Thus, while our results compare favorably with those of Salmeron et al., it is not clear that either approach will be a helpful guide to system hardening, mainly because hardening seems unlikely to be cost effective.

7. Conclusions and Directions for Future Research

In this paper, we developed a relatively simple, inexpensive, and viable method of identifying promising attack strategies. The impacts of our Max Line interdiction strategies for two sample transmission grids are comparable to interdiction strategies developed by Salmeron et al. (2004). However, our method and that developed by Salmeron et al. identify different sets of vulnerable transmission lines. Therefore, a single run of either method will likely not identify all critical vulnerabilities. Moreover, our results suggest that hardening transmission lines is not likely to be cost effective, since interdiction can cause substantial unmet demand even after significant hardening.

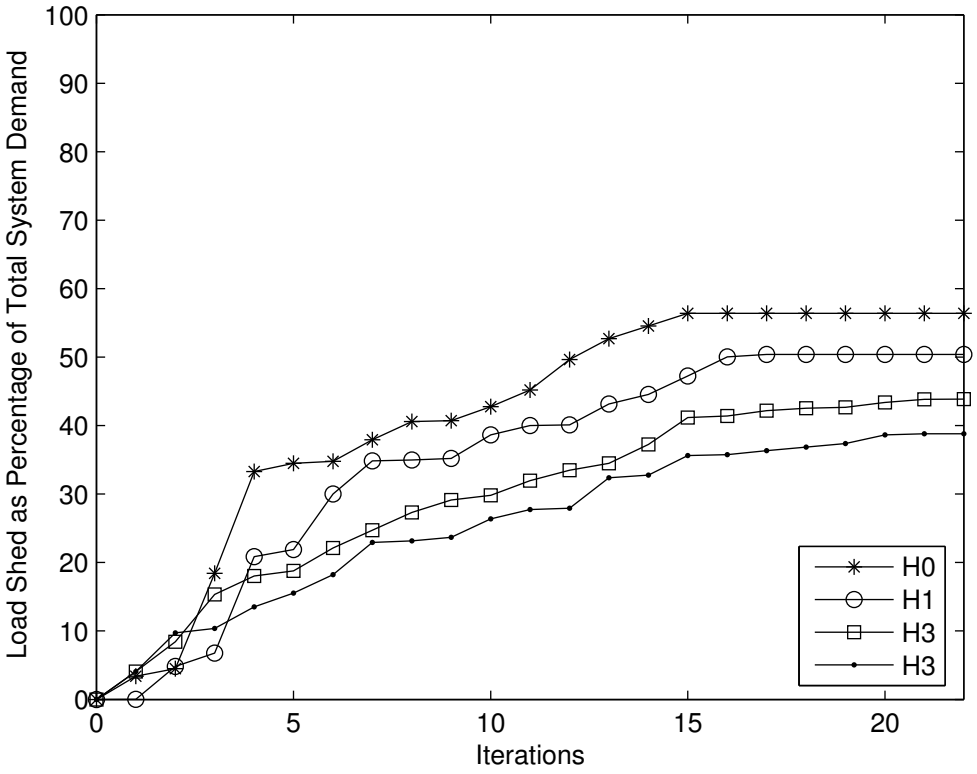


Figure 5: Interdiction strategies generated after hardening of the Two Area RTS-96.

Our work so far does have some important caveats. First, we considered transmission lines to be the only vulnerable components of a transmission system. Moreover, our interdiction and load-flow algorithms consider only power flows, and not the criticality of particular loads or demands.

In future research, this method could be extended to address other components of transmission systems, such as transformers (which would be represented as nodes rather than arcs). This is an important extension, since Zimmerman et al. (2005) note that transformers are especially difficult and time consuming to replace. It would also be desirable to extend the algorithm to identify interdiction strategies that may trigger cascading power failures. The possibility of cascading power failures was not considered in our algorithm, but could obviously amplify the effectiveness of line interdiction, as shown in the blackout of August 2003 (Electricity Consumers Resource Council, 2004). Finally, it would be helpful to adapt our algorithm to take into account the importance of different loads. In particular, Zimmerman et al. (2005) note that disrupting electrical supply to certain demand sectors (for example, transportation, or other types of critical infrastructure that depend on electricity) can have disproportionate impacts.

We also believe that the general approach outlined in this paper (the Max Line greedy interdiction algorithm) could be extended to identify critical components in other types of systems, such as structures (Schaefer and Bajpai, 2004, 2005; Bajpai and Schafer, 2003), water distribution systems (Michaud and Apostolakis, 2005), and ground transportation systems. Of course, the algorithm for re-optimizing load (in structures) or flow (in water or transportation systems) would be different from the load-flow algorithm used here for electricity transmission systems. However, we believe that the general approach of the Max Line algorithm could still be applied to such systems with reasonable results.

References

G. E. Apostolakis and D. M. Lemon, "A Screening Methodology for the Identification and Ranking of Infrastructure," Risk Analysis, Vol. 25, No. 2, 2005.

P. Bajpai and B. W. Schafer, "Progress Towards Structural Design of Unforeseen Catastrophic Events," ASME National Congress, Washington, D.C., 2003.

B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, "Critical Points and Transitions in an Electrical Power Transmission Model for Cascading Failure Blackout," CHAOS, Vol. 12, No. 4, 2002.

J. Chen, J. S. Thorp, and M. Parashar, "Analysis of Electric Power System Disturbance Data," Hawaii International Conference on System Sciences, Maui, HI, 2001.

Electricity Consumers Resource Council, "The Economic Impacts of the August 2003 Blackout," Washington, D.C., 2004.

B. C. Ezell, J. V. Farr, and I. Wiese, "Infrastructure Risk Analysis Model," Journal of Infrastructure Systems, Vol. 6, No. 3, 2000a.

B. C. Ezell, J. V. Farr, and I. Wiese, "Infrastructure Risk Analysis of Municipal Water Distribution System," Journal of Infrastructure Systems, Vol. 6, No. 3, 2000b.

B. C. Ezell, Y. Y. Haimes, and J. H. Lambert, "Cyber Attack to Water Utility Supervisory Control and Data Acquisition (SCADA) Systems," Military Operations Research, Vol. 6, No. 2, 2001.

G. L. Guzie, "Vulnerability Risk Assessment," U.S. Army Research Laboratory, 2000.

H. Liao, J. Apt, and S. Talukdar, "Phase Transitions in the Probability of Cascading Failures," Carnegie Mellon Electricity Industry Center, Working Paper 04-08, 2004.

Los Alamos National Laboratory, "Critical Infrastructure Protection Decision Support System (CIP/DSS) Project Overview," LA-UR-04-5319, Los Alamos, NM, 2004.

D. E. Michaud and G. E. Apostolakis, "Screening Vulnerabilities in Water-Supply Networks," unpublished technical paper, 2005.

L. Mili, Q. Qiu, and A. G. Phadke, "Risk Assessment of Catastrophic Failures in Electric Power Systems," International Journal of Critical Infrastructures, Vol. 1, No. 1, 2004.

North American Electric Reliability Council, "An Approach to Action for the Electricity Sector," Working Group Forum on Critical Infrastructure Protection, Princeton, NJ, 2001.

A. G. Phadke, "Hidden Failures in Electric Power Systems," International Journal of Critical Infrastructures, Vol. 1, No. 1, 2004.

Reliability Test System Task Force of the Application of Probability Methods Subcommittee, "The IEEE Reliability Test System – 1996," IEEE Transactions on Power Systems, Vol. 14, No. 3, 1999.

J. Salmeron, K. Wood and R. Baldick, "Analysis of Electric Grid Security Under Terrorist Threat," IEEE Transactions on Power Systems, Vol. 19, No. 2, 2004.

B. W. Schafer and P. Bajpai, "Stability Degradation and Redundancy in Damaged Structures," Annual Technical Session and Meeting, Structural Stability Research Council, Long Beach, CA, 2004.

B. W. Schafer and P. Bajpai, "Building Structural Safety Decision-Making for Severe Unforeseen Hazards," Proceeding of the 2005 NSF DMII Grantees Conference, Scottsdale, AZ, 2005.

R. Zimmerman, C. E. Restrepo, N. J. Dooskin, R. V. Hartwell, J. I. Miller, W. E. Remington, J. S. Simonoff, L. B. Lave, and R. E. Schuler, "Electricity Case: Main Report – Risk, Consequences, and Economic Accounting," Preliminary Report, 2005.

Acknowledgements

This material is based upon work supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number DAAD19-01-1-0502, the U.S. National Science Foundation under grant number ECS-0214369, and the Department of Homeland Security under grant number EMW-2004-GR-0112. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors. The authors would also like to acknowledge Prof. Ian Dobson of the Department of Electrical and Computer Engineering at the University of Wisconsin-Madison for his guidance and helpful contributions to this study.

Identification of Critical Locations across Multiple Infrastructures for Terrorist Actions

Sean A. Patterson and George E. Apostolakis*

Department of Nuclear Science and Engineering, and
Engineering Systems Division
Massachusetts Institute of Technology
Cambridge, MA 02139-4307, USA

Abstract

This paper discusses a possible approach to ranking geographic regions that can influence multiple infrastructures. Once ranked, decision makers can determine whether these regions are critical locations based on their susceptibility to terrorist acts. We identify these locations by calculating a value for a geographic region which represents the combined values to the decision makers of all the infrastructures crossing through that region. These values, as well as the size of the geographic regions, are conditional on a minor destructive threat of a given size, e.g,. a bomb that can affect objects within 15 feet of it.

This approach first requires an assessment of the users of the system. During this assessment, each user is assigned a performance index (PI) based on the disutility of the loss of each infrastructure's resource via multi-attribute utility theory (MAUT). A Monte Carlo network analysis is then performed to develop importance measures (IM) for the elements of each infrastructure for their ability to service each user. We combine the IMs with the user PIs to a value that we call valued worth (VW) for each infrastructure's elements independently. Then we use spatial analysis techniques within a Geographic Information System (GIS) to combine the VWs of each infrastructure's elements in a geographic area, conditional on the threat, into a total value we call geographic valued worth (GVW). The GVW is graphically displayed in the GIS system in a color scheme that shows the numerical ranking of these geographic areas. The map and rankings are then submitted to the decision makers to better allocate anti-terrorism resources.

A case study of this methodology is preformed on the Massachusetts Institute of Technology's (MIT) campus. The results of the study show how the methodology can bring attention to areas that may be ignored through individual infrastructure analysis. The intersections of major infrastructures on the campus prove to be of the most importance to the stakeholders of the campus.

Keywords: Infrastructures, Networks, Terrorism, Risk Analysis

* Corresponding author. Office: Room 24-221; Email address: apostola@mit.edu; fax: +1-617-258-8863

† Relative variance is a technique used in Monte Carlo methods to reduce the error of the results. It helps determine a sufficient sample size required to increase confidence in the simulation results.

1. Introduction

After the September 11th, 2001 attack, the U.S. Government created a list of infrastructures considered to be critical to the United States. These critical infrastructures are, by default, potential targets (Office of Homeland Security, 2002). These infrastructures are complex and interdependent. This massive intricacy poses a financial allocation dilemma for government and industry. Previous reports such as the one issued by the National Research Council (National Research Council, 2002) offer a large number of recommendations to protect these infrastructures. The cost/risk-reduction of all of these measures is not evident. Implementing all of these recommendations would impose a large financial burden on governments to implement all proposed measures. A screening methodology is therefore needed to

determine the allocation of financial resources.

The problem of screening for terrorist vulnerabilities on a critical infrastructure as it impacts society is complex. High-level screening can give an indication as to how resources should be allocated in order to better protect society. One example of this high-level screening was presented by Paté-Cornell and Guikema (2002). This model is characterized as "overarching," i.e., it does not go into the analysis on the physical networks. Garrick et. al. (Garrick et. al., 2004) recommend that a scenario-based methodology known as Probabilistic Risk Assessment (PRA) be used to identify, quantify, and manage terrorist threats. Apostolakis and Lemon (2005) propose the use of PRA to screen terrorism scenarios on infrastructures.

The Apostolakis and Lemon methodology combines multiattribute utility theory (MAUT) and PRA and is demonstrated on the campus of the Massachusetts Institute of Technology (MIT). It determines the disutility of users caused by the loss of each of the infrastructure networks under analysis. Cut set analysis is then performed on the networks and each cut set (consisting of nodes and arcs) is assigned values based on the amount of disutility it creates for its users. Analysis is then performed on the susceptibility of the cut sets to attack and all node/arcs are ranked according to their values. This model uses the physical networks to screen for vulnerabilities. It is important to note that nodes (e.g., manholes) that had different infrastructures running through them were given the same node name and therefore geographically common nodes were identified by visual inspection by the authors.

Michaud and Apostolakis (2005) propose that cut set analysis is too stringent in real systems. They stress that sets of node/arc losses that do not fully limit flow can restrict resources to the point that it is virtually a cut set. The case study they use involves the water-supply network of a medium-size city. Due to the capacity limitation, the users do not have unlimited access to water. These limitations may, in effect, cut a user off from the network despite the user' physical positive connectivity to the resource.

Different infrastructures coincide geographically in a complex manner. When the geographic locations of infrastructures are plotted on a map, it can be seen that the infrastructures physically overlap or come spatially very close. Thus, an intentional attack on one infrastructure, specifically with a bomb, will more often than not affect other overlapping or nearby infrastructures.

We propose a screening methodology for identification and prioritization of geographic regions. This is accomplished through the combination of the framework of Apostolakis and Lemon, Monte Carlo network analysis methods, and geographic analysis methods. The analysis is conditional on a destructive threat, e.g., a bomb. Though we develop a numerical rank for the infrastructure elements from this screening, we present it in a graphical form that can show geographic concentrations of elements that cause large increases in risk for the given threat. The ranking is developed through MAUT, which allows us to develop our rankings using the stakeholder and decision maker values.

This paper is arranged by first covering an overview of the methodologies our predecessors at MIT have done followed by a section on our contribution to these methods. Then we present an in-depth methodology overview using examples from our case study. Following this, we present the results of our case study. We end with a few conclusions about the proposed methodology.

2. Predecessor Works

Two predecessor works have been completed at MIT to define a new approach to infrastructure analysis. Much of these works is the foundation of the proposed methodology of this paper.

*2.1. A Screening Methodology for the Identification and Ranking of Infrastructure Vulnerabilities due to Terrorism (*Apostolakis and Lemon, 2005*)*

These authors developed a screening methodology to prioritize critical locations of infrastructures for a minor terrorist attack. They note that the national infrastructures are owned by several stakeholders. Therefore, to include the values of these stakeholders they use MAUT to treat risk as a multiattribute concept and give a consistent basis for the ranking of vulnerabilities. They also assume a minor threat defined to be a single point attack against one or more infrastructures resulting in minimal restoration.

Vulnerability is defined as the "manifestation of the inherent states of the system (e.g. physical, technical, organizational, cultural) that can be exploited by an adversary to harm or damage the system." (Haimes and Horowitz, 2004) A *threat* is "a potential intent to cause harm or damage to the system by adversely changing its states" (Haimes and Horowitz, 2004). This threat is an initiating event in PRA language (Garrick, et al, 2004). Infrastructures were built for efficiency and convenience, and are therefore

are open and accessible particularly during malevolent attack (Haimes and Horowitz, 2004). Therefore, "the concept of vulnerability includes both a measure of how accessible to terrorism a particular target is and the system-damaging sequence of events that may be initiated after this target is attack. The evaluation of the threat is usually left to the intelligence agencies. The identification of vulnerabilities given a threat is a technical problem." (Apostolakis and Lemon, 2005) Their screening methodology focuses on the identification of critical locations. These critical locations are "part of the vulnerabilities. They are defined as geographic points that are susceptible to attacks." Critical locations are not limited to a single infrastructure, but may affect multiple infrastructures at the same location (e.g., a manhole with access to water and gas). We will use these same definitions throughout our paper.

The first step of the methodology is the selection of the assets to be protected. In the case study, they chose the electric, domestic water, and natural gas systems of the MIT campus, and therefore they determined what campus facilities needed an uninterrupted supply of these recourses. The next step is the identification of scenarios initiated by a minor threat that would lead to interruption of the services. To do so, the relevant infrastructures are modeled so that minimal cut set (mcs) analysis is easy to perform. A minimal cut set is a set of events that assure the interruption of supply to a user. All the events in a mcs are required for the interruption. The authors used a network diagraph to model each of the three networks. Supply and user nodes where identified as well as the network vertices and arcs. Vertices which had a common geographic location with other infrastructure vertices are labeled the same name. This identification of geographically common intersections was done by visual inspection.

The mcs must be assigned a value in order to perform a ranking. Apostolakis and Lemon argue that the prioritization should be based on the expected value to the decision maker of the consequences of the vulnerabilities. Such a scheme would require an evaluation of the conditional probability that the terrorists will actually attack a given mcs successfully, something which is inherently difficult to evaluate. The authors, therefore, separate the vulnerability's value from the conditional probability of a successful attack. However, they do provide additional information to the decision maker regarding the degree to which a potential target is accessible, i.e., susceptibility judgments.

As stated above, Apostolakis and Lemon use MAUT to assess the value of the mcs to the decision maker. A performance index (PI) is calculated for each mcs. The PI is shown in eq. (1) and is the sum of the weights of individual performance measures (PMs) multiplied by the disutility the loss of an infrastructure causes the user in the context of the respective PM.

$$PI_{jk} = \sum_{i}^{K_{pm}} w_i d_{ijk} \qquad (1)$$

where:

PI_{jk} is the performance index for user *j* for loss of infrastructure *k*

w_i is the weight of the performance measure *i*

d_{ijk} is the disutility of performance measure *i* for user *j* for loss of infrastructure *k*

K_{pm} is the number of performance measures

When $PI_A > PI_B$ the decision maker assesses case A to cause more disutility than case B. Examples of these PM from their MIT case study are: impact on people and impact on external public image. The PMs are developed systematically using a value tree, which is representative of the concerns of the stakeholders and is a hierarchal approach to structuring underlying PMs to overall objectives (Gregory and Keeney, 1995; Clemen, 1996). The value tree from the MIT case study, which we will also use in our case study, is shown in . This value tree is based on a value tree that had been developed in an independent deliberative process that the MIT Department of Facilities had held with a group of MIT stakeholders (Karydas and Gifun, 2002). Figure 1

The relative weights of the PMs are also produced in the deliberative process. In addition, constructed scales are developed for each PM so that the decision maker can assess how an event affects a user. This event in cases involving infrastructures is the loss of the supply to a user for the infrastructure. The assessment of what level of the constructed scale is affected by an event is left to the decision maker. A constructed scale level is picked for all PMs of each user for each event. An example of a constructed scale is in Table 1.

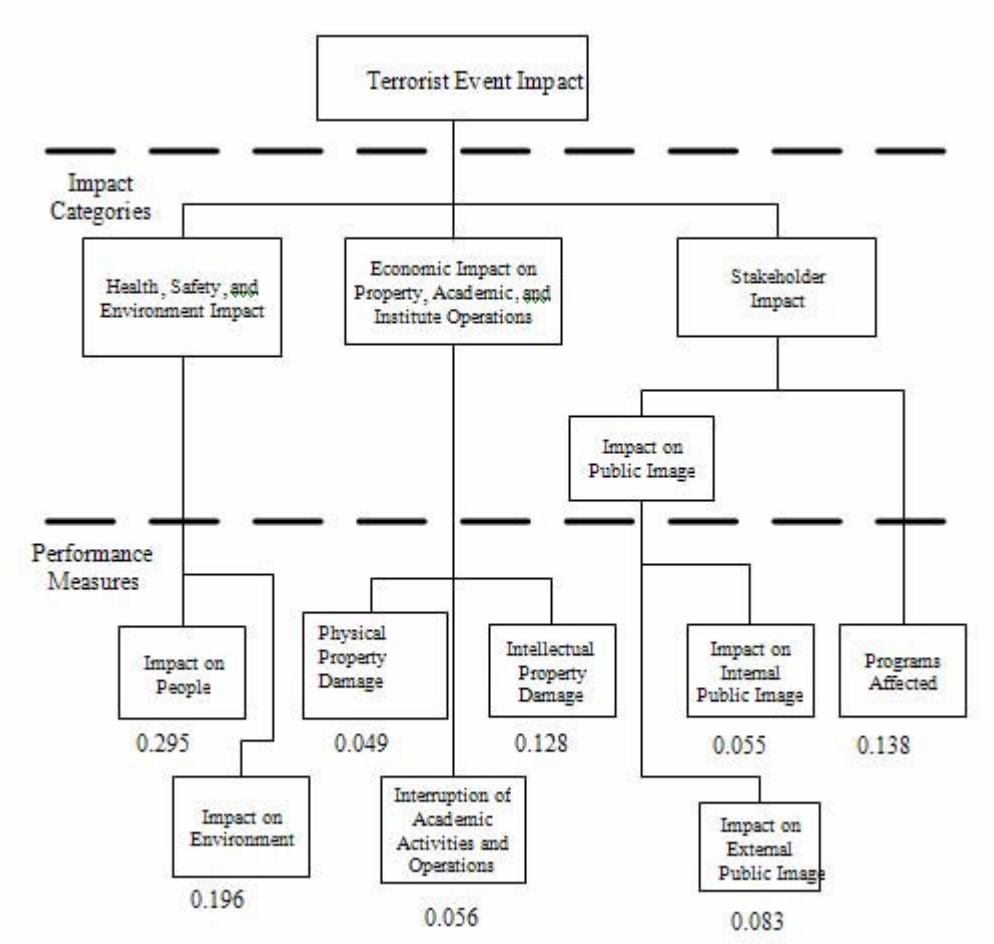


Figure 1: Value tree and weights for the MIT case study.

Level	Description	Disutility
3	Fatality or Lethal Exposure, e.g., Roof Collapse, Falling Brick, Inhalation of as	1.00
2	Major Exposure with Long Term Effects, e.g., Lead Poisoning	0.46
1	Minor Injury or Exposure, e.g., Broken Arm, Laceration	0.05
0	No personal injury	0.00

Table 1: Constructed Scale for Impact on People

Once the decision maker has assessed the PI for all users for all events, the mcs PIs can be calculated. Apostolakis and Lemon use Equation (2) to evaluate these PIs.

$$PI_y = \sum_j \sum_k (mcs_y^{jk} PI_{jk}) \qquad (2)$$

where:

PI_y is the performance index for mcs *y*

mcs_{jky} is a Boolean operator (1 when the mcs *y* impacts the

user-infrastructure combination *jk*, and 0 otherwise)

Equation (2) sums all the disutilities that a minimal cut set creates to the users of each infrastructure. The result is a PI ranking for all of the mcs. It is evident that, when mcs are common among two or more infrastructures, these mcs are of the highest value to the users. In the MIT study, this mainly occurred in the low–order mcs, i.e., mcs that involved one or two elements only.

To address the issue of vulnerability, Apostolakis and Lemon created a scheme to combine the values (PI of the mcs) with susceptibility for each mcs. Their process of determining susceptibility is subjective. Using Table 2, they assign a susceptibly level to each mcs. The authors then combine the susceptibility evaluation with the PI ranking to produce a ranking of the vulnerabilities in a categorical manner presented in Table 3.

Level	Description (examples)
Extreme	Completely open, no controls, no barriers
High	Unlocked, non-complex barriers (door or access panel)
Moderate	Complex barrier, security patrols, video surveillance
Low	Secure area, locked, complex closure
Very Low	Guarded, secure area, locked, alarmed, complex closure
Zero	Completely secure, inaccessible

Table 2: Susceptibility categories

Vulnerability	Description
Red	This category represents a severe vulnerability in the infrastructure. It is reserved for the most critical locations that are highly susceptible to attack. Red vulnerabilities are those requiring the most immediate attention.
Orange	This category represents the second priority for counter-terrorism efforts. These locations are generally moderately to extremely valuable and moderately to extremely susceptible.
Yellow	This category represents the third priority for counter terrorism efforts. These locations are normally less vulnerable because they are either less susceptible or less valuable than the terrorist desires.
Blue	This category represents the fourth priority for counter terrorism efforts.
Green	This is the final category for action. It gathers all locations not included in the more severe cases, typically those that are low (and below) on the susceptibility scale and low (and below) on the value scale. It is recognized that constrained fiscal resources is likely to limit efforts in this category, but it should not be ignored.

Table 3: Vulnerability Categories

This color coding gives the decision maker a good qualitative judgment based on quantitative facts to determine how best to allocate money to protect the his interests and those of the stakeholders.

A major finding of the Apostolakis and Lemon work was that the mcs with the highest vulnerability was a manhole through which the three infrastructures (electric, natural gas, and water) pass. The mcs for the loss of all three infrastructures to a user was a single node (the manhole). This single node caused a large PI in the analysis. Combining this high PI with the finding that the manhole was very accessible and thus at an extreme susceptibility level (Table 2), the authors concluded that this single node belonged to the red vulnerability category (Table 3). This is a critical location because it has a high vulnerability due to geographic coincidence of multiple infrastructures and thus depends on the geographic layout of the infrastructures.

2.2. Screening Vulnerabilities in a Water-Supply Network

Another work done at MIT (Michaud and Apostolakis, 2005) developed another methodology using the latter work as a basis. This work was specifically developed for a water-supply network but suggests that it may be applicable to other infrastructure types. Therefore, unlike Apostolakis and Lemon, it was not a multi-infrastructure analysis. The goal of the research was to develop a screening methodology for water-supply network vulnerabilities to terrorism. This research specifically took into account capacities and repair, and was calculated through a GIS program. Michaud and Apostolakis added and changed several things from the Apostolakis and Lemon methodology to accomplish this goal, these areas will be pointed our where appropriate.

The infrastructure is first modeled with a Geographic Information System, ESRI ARCGIS in this case. Geographic Information Systems (GIS) are programs that display geospatial information stored in a database in graphical form. Their open architecture allows users to code analysis programs based on spatial requirements. This is in contrast to the digraph models used by Apostolakis and Lemon. The network must have the capacity and repair time of the network elements included as attributes of the element in a GIS database. Flow directions are also setup in the network so looping cannot occur.

Instead of a minimal-cut-set network analysis, this methodology looks for a loss of capacity to a user vice catastrophic loss of a system. Single failures where assumed for each arc and node connecting each user to a source. When the program assumed a failure of an arc/node and detected that the user could not receive its full required water supply, it picked a constructed scale level based on the new supply capacity caused by the failure of the respective network element. The constructed scale level picked was a function of the capacity loss to the user as well as the ability/time of workers to repair the element. The PI of the user is then calculated using Equation (1).

2.3. This Paper's Contributions

The research presented in this paper uses parts of the above two methods as a starting point, but seeks to expand and change several areas. We take the broad context of the above methodologies by combining network analysis with the PIs of users for the loss of those networks. Therefore, we accept the use of MAUT as the main vehicle to calculate values of the individual users. We will keep the concepts of disutility, value tree, PMs, and stakeholders since they are an effective way to screen these vulnerabilities. We briefly describe our expansions and modifications in the following paragraphs.

Our analysis takes place on a much grander scale than our predecessor works. We analyze 133 users and 5 infrastructures. Because of this, we did not want to assess individually each constructed scale per user, as Apostolakis and Lemon did. We also did not want to dynamically pick the constructed scale within our program like Michaud and Apostolakis. This was due to the required computation time of our network analysis, discussed later. Instead, we will present a method to group users and diversify them using GIS attribute data about the buildings, e.g., the number of people residing in a building, floor space of the building, etc. This is done programmatically before the network analysis and provides diversified users without much input from the decision maker.

We agree with the divergence from the minimal-cut-set analysis that Michaud and Apostolakis performed. We too will diverge but not base our analysis on capacity and single component failures. We will present a method to develop importance measures using Monte Carlo network analysis that will give us answers about what the failure of elements means to the system.

Like Apostolakis and Lemon, we develop a multi-infrastructure analysis. However, we do not identify common nodes by inspection followed by a minimal-cut-set analysis. We instead perform a network analysis on all infrastructures independently and then, through GIS algorithms, we find geographically coincident and even spatially close nodes/arc, e.g., parallel pipes within a certain distance from each other. Doing this "intersection analysis" after the network analysis allows us to easily change the "intersection distance," i.e., how close different elements must be to be considered spatially coincident.

Like Michaud and Apostolakis, we use GIS as a tool and programming platform. We too present our findings in a graphical display. However we will develop a grid which will determine the increase in risk to society for geographic regions. These regions (grid spaces) are conditional on a minor threat (ex. a single bomb attack) and have a value based on their ability to increase in risk to society for minor threat of a given radius of influence. This radius is the range of a destructive threat, e.g. a bomb, which destroys elements of infrastructures within the radius. For example, a bomb that can affect a 5 ft radius, any infrastructures within 10 feet of each other must be analyzed by a concurrent initiating event. The map we later present is therefore conditional on the type and size of the threat.

The rest of this paper will describe how we calculate the values of these geographic regions which will be displayed in a conditional map.

3. Methodology Overview

Here we present an overview of the proposed methodology for screening critical locations using GIS.

3.1. Performance Index Assessment

First we identify the infrastructures of interest. These infrastructures must be in a GIS database or convertible file format. Each infrastructure should have at least one supply within the scope of the analysis. We use the same equation, Equation (1), as Apostolakis and Lemon. Before we go on to assess constructed scales, we pause to reflect on our task of assessing hundreds of users for many events. In order to create a manageable PI assessment for our 133 users and all five events, we decided that we should not individually analyze each user for its PI. Rather we created a user hierarchy by grouping users into Macro-user Groups (MGs) based on the main function of the building. For example, dorms would be in the residential MG and a building of classrooms would be in the academic and research MG. With this hierarchy established, we now only assess the PMs of users by group, i.e., all the users within an MG have the same constructed scale level picked for a given event. By grouping the users into their respective MG, we reduce the number of decisions that must be made by the decision maker since the users within the MG are dependent on the MG itself not the decision maker's individual assessment.

Due to the MG scheme presented above, so far all, the users in the same MG have the same PI for a given event, i.e., loss of an infrastructure. Obviously, it is not the case that all residential MGs have the same value to the decision maker for the same event. With the use of GIS and the addition of the MG scheme we can add another layer of diversification to all users. To do this, we apply a natural scale to the constructed scale. A natural scale, such as the amount of classroom square footage in a building, can be multiplied by the constructed scale to yield a weighted disutility for a certain PM. We call this natural scale the weighting function since it weights the impact of the constructed scale by some data. Therefore "d" in Equation (1) is now called the weighted disutility and is equal to the unweighted disutility from the constructed scale times the value from the weighting function. The key here is to use data that are available within GIS or some accessible database. By doing this the scaling process the weighting function performs on the PMs is not a decision the decision maker makes for each user. Rather it is a mathematical calculation made on the data of the user. The mathematical function itself is set during the frame working process that the stakeholders perform. Once the function provides positive consistency checks, as is the case for all of the weighting in the value tree, it does not change during the actual PI assessment process, thereby eliminating individual assessment by the decision maker.

As said, this weighting function is established during the frame working process by the stakeholders for PMs where they believe it is applicable and where it leads to consistent results. A PM can have a weighting function based on one type of data or a combined data. The weighting function must however scale this data so that output is greater than zero and saturates at one. The functions can be anything from linear to non-linear as long as the PIs are consistent.

Let us look at this concept in an example. In Table 4, we present two dorm buildings both within the residential MG. One building houses 500 people and the other 50 people. We calculate the PI for the scenario where the only PM impacted is the "impact on people" whose constructed scale is shown in Table 5. We assume that an event leads to an impact assessed as level 3. Without a weighting function the PI is calculated by using Equation (1) and the value tree from Figure 1. Therefore each building has a PI of 0.295. Thus, before the weighting function is applied, the two dorms, by virtue of being in the same MG, have the same PI.

Now, we assume the same arbitrary event thus choosing the level 3 impact on the "impact on people" PM and keeping the other PMs at level 0, but this time we apply a weighting function (Figure 2) based on the number of people affected. In this case, the two dorms have different PIs for the same event. What this really means is that by using a population weighting function, the 500 person dorm evokes more disutility because more people are affected.

	Population Data	Old Total PI	Weighting Function	New Total PI
Dorm A	500 people	0.295	.99	.292
Dorm B	50 people	0.295	.63	.186

Table 4: Comparison of using and not using a weighting function

Level	Description	Unweighted Disutility
3	Fatality or Lethal Exposure	1.00
2	Major Exposure with Long Term Effects; Loss of jobs	0.46
1	Minor Injury or Exposure; Significant Employment interruption	0.05
0	No personal injury or job loss	0.00

Table 5: Constructed Scale for the PM "impact on people."

These functions are not required to be used for all PMs or MGs. It is even possible to setup constructed scales to handle this data based weighting during the initial framework process. For example, we can see that currently the constructed scale in Table 5 has levels based on the intensity of the impact on the PM. Instead we could create three PMs: Minor impact on people, Major impact on people, and Fatalities. These three PMs would then have levels where the descriptors would be the number of people impacted in the context of the PM, i.e., the lower the level the fewer people impacted. However, trying to keep this data scaling and adjusting the constructed scale to accommodate this data will most likely lead to an increase in the number of PMs, as in the example just given. In our case study, we wanted to keep the constructed scales and PMs set up by Apostolakis and Lemon thus we use a weighting function on their established constructed scales.

No matter what way PIs are assessed, with or without MGs and weighting functions, consistency must be established. It may be found that using or not using our suggestions can either more easily or less easily establish consistent PIs. Each decision analysis process is unique, thus it is hard to say that the MG scheme and weighting functions will help other PI assessment processes. However, we found it useful in our case study and we will use both the MG and weighting function schemes.

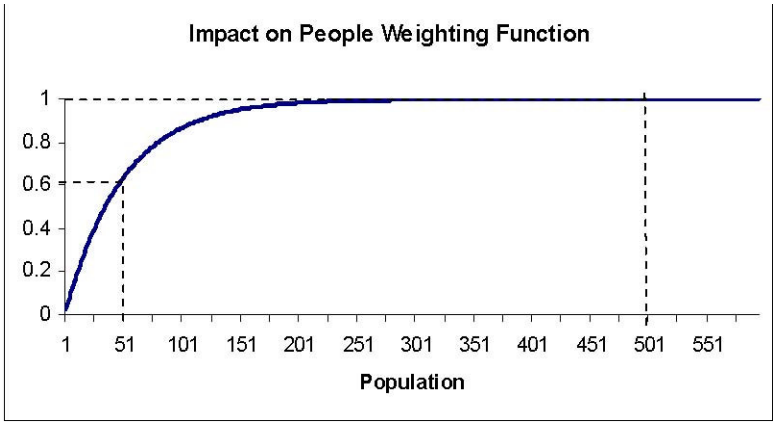


Figure 2: Impact on People Weighting Function

3.2. Importance Measures

It has been stated that minimal-cut-set analysis becomes obsolete as the network becomes larger and highly distributed. (Billinton, 1992; Marseguerra and Zio, 2002) Cut-set analysis is also computationally intense. To evaluate the network elements we develop importance measures (IMs). These measures are quantifications of how the availability or unavailability of the elements affects the network. This section describes several IMs that will be calculated using Monte Carlo simulations.

There are several importance measures that have been developed in the literature. We will calculate four of them: Fussell-Vesely (FV) Eq. (3), Birnbaum Eq. (4), Risk Achievement Worth (RAW) Eq. (5), and Risk Reduction Worth (RRW) Eq. (6). FV and RRW are related mathematically. Birnbaum completely depends on network structure. (Cheok, Parry, and Sherry, 1997). Descriptions of what the each IM means is given below their respective equation.

$$\mathrm{FV}_{ykj} = \frac{\mathrm{U}_{kj} - \mathrm{U}^{-}_{ykj}}{\mathrm{U}_{kj}} = 1 - \frac{1}{\mathrm{RRW}_{ykj}} \quad (3)$$

Fussell-Vesely (FV) describes the maximum fractional decrease in risk to the user *j* for infrastructure *k* when element *y* is made always available.

$$B_{ykj} = U^{+}_{ykj} - U^{-}_{kyj} \quad (4)$$

Birnbaum describes the maximum change in risk to user *j* for infrastructure *k* when element y switches from available to unavailable.

$$\mathrm{RAW}_{ykj} = \frac{\mathrm{U}^{+}_{ykj}}{\mathrm{U}_{kj}} \quad (5)$$

Risk Achievement Worth (RAW) describes the ratio of risk to user *j* for infrastructure *k* to the risk to the user when element *k* is always unavailable.

$$\mathrm{RRW}_{yku} = \frac{\mathrm{U}_{kj}}{\mathrm{U}^{-}_{ykj}} \quad (6)$$

Risk Reduction Worth (RRW) describes the ratio of risk to user *j* for infrastructure *k* to the risk to the user when element *k* is always available.

(Zio, Podofillini, and Zille, 2006)
where:

U_{kj} percent of simulations in which there is no path connecting the user *j* to an infrastructure *k* source

U^{+}_{ykj} percent of simulations in which element *y* of infrastructure *k is failed* and there is no path connecting the user *j* to an infrastructure *k* source

U^{-}_{ykj} percent of simulations in which element *y* of infrastructure *k is not failed* and there is no path connecting the user *j* to an infrastructure *k* source

The IMs are calculated using U_{kj}, $U^{+}_{ykj,}$ U^{-}_{ykj}. Since we will be performing multiple Monte Carlo simulations, described below, to calculate these IMs these values are calculated by keeping counters for each element as well as the system and dividing them appropriately. For the system there is a total counter U_{total}, which is also the number of simulations run for any given user-infrastructure combination. There is a counter $U_{system--fail}$ which tracks the number of simulations where the user is not connected to a supply. For each element of the infrastructure there are four counters, two when it is failed and two when it is working. The counter $U_{k\text{-}fail}$ tracks the number of times the element is failed and the counter $U_{k\text{-}fail\text{-}system\text{-}fail}$ tracks whether the element is failed while the system is failed. There is counter $U_{k\text{-}working}$ which tracks the number of times the element is not failed ($1-U_{k\text{-}fail}$) and $U_{k\text{-}working\text{-}system\text{-}fail}$ for when the element is working but the system is failed. U_{kj}, $U^{+}_{ykj,}$ U^{-}_{ykj} are therefore given by the following Equations (7) through (9):

$$U_{kj} = \frac{U_{system-fail}}{U_{total}} \quad (7) \qquad U^{+}_{yjk} = \frac{U_{k-fail-system-fail}}{U_{k-fail}} \quad (8) \qquad U^{-}_{yjk} = \frac{U_{k-working-system-fail}}{U_{k-working}} \quad (9)$$

(Zio, Podofillini, and Zille, 2006)

In our case study (presented later), we calculated all four IMs listed above, however the first thing we noticed about our final rankings was that two the results using two IMs yielded the same ranking. These two (RRW and FV) gave us a sanity check since, as shown in eq. (3), they are mathematically related. The results based on RAW also gave a ranking which was similar to the results based on RRW and FV. Thus, in order to give the decision makers an easily assessable number set we will present our equations and findings using the RAW. We also note that RAW is "commonly used as an intuitive measure of margin provided by the component." (Cheok, Parry, and Sherry, 1997) However, in the equations below, RAW can be replaced by any of the other three importance measures, as long as that IM is consistently used for all equations, in order to get a different insight into the system.

3.3. Network Analysis

In order to develop the above counters and thereby get the IMs (RAW) of the elements, simulations must be preformed for each user *j* per each infrastructure *k*. These simulations are random network states created through Monte Carlo (MC) trials. We present a time-independent simulation. This means that there needs to be a probability associated with the random failures that are induced in each independent simulation. This value could be thought of as a probability of a successful attack, i.e., the probability that, conditional on the threat, the threat would induce a failure; e.g., a suitcase bomb placed atop a manhole would fail the component(s) and/or pipe(s) of an infrastructure(s) inside of the manhole. This probability can be associated to a target's susceptibility to the conditional threat. If we run a Monte Carlo simulation where all components have the same probably of failure, i.e., all elements have an equal likelihood of failing from the same threat, then the random failures which occur in a reliable simulation will yield the same ranking (presented later) even if the probably is changed (keeping the same value for all components) and only yield different importance measure values. This means that as a first cut we can get a ranking of components without expert analysis as to the susceptibility of the elements under analysis. Ignoring susceptibility at this point means that the rankings we develop later do not suggest an element is vulnerable

or not vulnerable, rather the elements pose a lesser or greater risk to the network due to their unavailability.

In future work, with expert analysis we could assign the proper probability of a successful attack to each network element independently. This does not mean the probability of an attack (ability of a group to desire to, obtain materials, and carry out an attack) needs to be assessed. We define the probability of a successful attack only by the ability of the threat (e.g., bomb) to destroy a target. It is the susceptibility of a target to a conditional threat, not the probability of the threat to be created. For example expert analysis would most likely render pipes under a street less probable to failure than pipes passing through an access point (manhole). By doing this the final rankings will be representative of the vulnerabilities since the simulations will now be based on the susceptibilities of the elements. If this is done care should be taken to realize that if the susceptibilities change for any reason, e.g. countermeasure are installed, all simulations must be run again.

Each simulation is one random network state. With a given user *j* and a single infrastructure *k*, we pick a random number, R_{ky}, for each network element *y* of infrastructure *k*. For each element, the random number R_{ky} is compared to the probability of a successful attack, PF, described above. If R_{ky} <PF, then the element is considered failed for the current simulation. Once all network elements have been appropriately failed the simulation state is considered set. The GIS program then uses an internal algorithm to check whether there is a path that connects user *j* to a supply node of infrastructure *k*. An existent path implies a working system and a non-existent path implies a failed system.

Once this path is checked the simulation is complete, we then proceed to update all of the counters (U_{total}, $U_{system\text{-}fail}$, $U_{k\text{-}fail,}$ $U_{k\text{-}fail\text{-}system\text{-}fail,}$ $U_{k\text{-}working,}$ $U_{k\text{-}working\text{-}system\text{-}fail}$) once, where appropriate. The state of the system is then reset so all elements are available. A new MC trail is preformed with new random numbers and again updating the counters appropriately. This process continues until all MC trials, N, have been performed for current user *j* and infrastructure *k*.

We then use the counters from the N trials to calculate the RAW for each element of the infrastructure *k.* Then we multiply the RAW for each element by the PI of user *j* given infrastructure *k.* This multiplication scales the RAW by the amount of disutility (PI) the loss of the infrastructure *k* creates for user *j*, i.e., the amount of disutility user *j* experiences because no path exists from any supply of infrastructure *k* to the user *j.* We therefore term the resultant calculation "worth" and describe it in Equation (10). Equation (10) also points out the addition of our weighted disutility versus equation (1) without the weighting function scheme.

$$W_{ykj} = RAW_{ykj} * PI_{jk} = RAW_{ykj} * \sum_{i}^{K_{pm}} w_i d_{ijk} \qquad (10)$$

where:

w_i is the weight of PM *i* from the value tree

d_{ijk} is weighted disutility of PM *i* for user *j* for loss of infrastructure *k*

W_{jky} is the worth of element *y* of infrastructure *k* for user *j*

PI_{jk} is the performance index for user *j* for loss of infrastructure *k*

RAW_{ky} is the importance measure RAW of element *y* of infrastructure *k* for user *j*

After all of the worths have been calculated for each element of infrastructure *k* we then reset the MC trial counter and perform the above simulations and calculations for another user *j*+1 for the same infrastructure *k*. This is done for all users that are serviced by infrastructure *k*. Once complete, we do a summation of all of the worths for each element for all users infrastructure *k* services. This summation assumes independence between the users just like we assume independence between PMs in Equation (1). This requires that if, for example, user *a* loses service from an infrastructure, the loss to user *a* alone cannot affect user *b's* service. User *b* service may only be affected by the network itself. Thus for user *b* to lose service, the infrastructure must have the appropriate components unavailable so that there is no path from a supply to user *b*; user *b* cannot simply lose an infrastructure just because user *a* lost service, i.e. only the network itself can affect a user; users cannot affect other users. We call the result of this summation "valued worth" and it is given by Equation (11).

$$VW_{yk} = \sum_{j} W_{ykj} = \sum_{j} \left[RAW_{ykj} * \sum_{i}^{K_{pm}} w_i d_{ijk} \right] \quad (11)$$

where:

VW_{yk} is the valued worth of element *y* of infrastructure *k*

W_{ykj} is the worth of element *y* of infrastructure *k* for user *j*

After the valued worth is calculated for each element of infrastructure *k*, we then reset the MC trials and advance to infrastructure *k*+1. All of the above calculations and MC simulations are preformed for each user serviced by infrastructure *k*+1, i.e., run N MC trails for each user tracking the counters and then calculating the worth of each element; once all users have been simulated N times then the valued worth of all elements of infrastructure *k*+1 is calculated.

The valued worths of a given infrastructure can be ranked and displayed in conditional risk maps. These maps give the decision makers an independent view of each infrastructure. The valued worths represent RAW scaled by the potential disutility they evoke to all users of their respective infrastructure. The higher the valued worth, the higher the probability that the unavailability of the network element will fail the network to the users it services with some disutility (PI) potentially caused to those users. It is important to note here that these values calculated through RAW do not give a yes/no answer. As said, the values represent the potential of an element's unavailability to fail the system and cause a certain amount of disutility; the failure of a high value item may or may not lead to a user(s) losing service to an infrastructure.

3.4. Combining Multiple Infrastructures

Once the valued worth for all elements of all infrastructures have been calculated we proceed to the full power of GIS, using spatial analysis, to develop our final results. First, we must develop a generic grid to be laid across the map of all the infrastructures. The side of each grid space is the size of the threat's radius of influence. We use a hexagonal close packed grid across the entire region of analysis. A generic grid is displayed in Figure 3. Figure 4 shows a grid laid over three infrastructures.

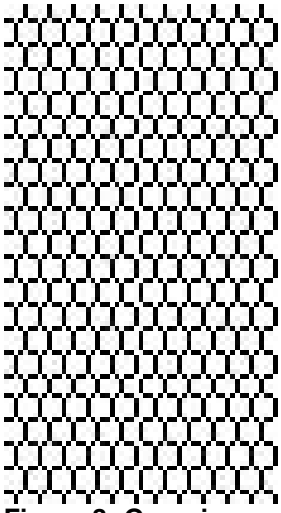


Figure 3: Generic hexagonal grid with a 5-meter radius of influence - zoomed in.

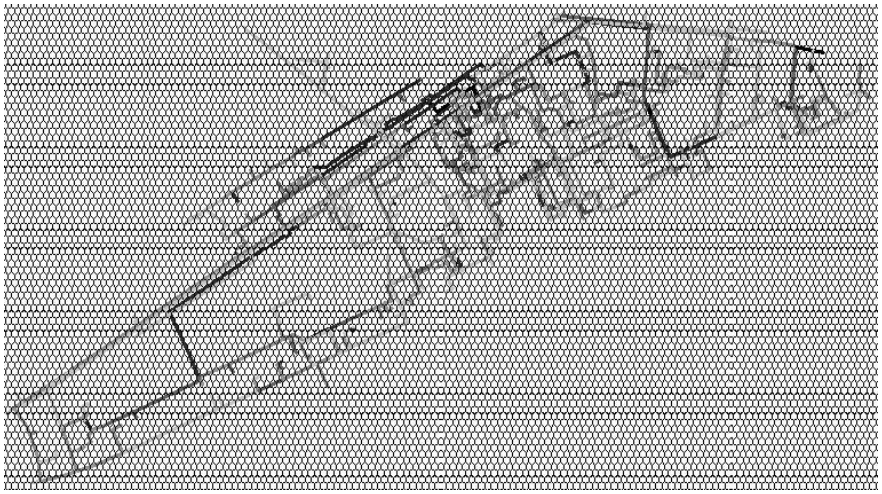


Figure 4: Generic hexagonal grid with a 5-meter radius of influence laid over three infrastructures.

We use an internal GIS function to first take the maximum valued worth of all elements of the same infrastructure that pass through or are located within each hexagon. We then sum the maximum valued worth elements from each infrastructure for each hexagon. Since the valued worths are now only relevant to their geographic location we call the result of this process the "geographic valued worth" (GVW). Equation (12) describes this process.

$$GVW_{xz} = \sum_{k} \max(VW_{y_{xz}k}) = \sum_{k} \max\left[\sum_{j} \left[RAW_{y_{xz}kj} * \sum_{i}^{K_{pm}} w_i d_{ijk} \right] \right] \quad (12)$$

where:

GVW_{xz} is the geographic valued worth of the grid space at coordinates (x,z)

$Max(VW_{y(xz)k})$ is the maximum valued worth element *y* out of all the elements of infrastructure *k* that pass through grid element (x,z)

Unlike the valued worth, GVW is a multi-infrastructure value. GVW is always conditional on the threat and thus the radius of influence. The GVW can be ranked in tabular format but the most intuitive presentation is through a color ramp of the GVW values and displaying them in a GVW conditional risk map. The color ramps used in our figures are grayscale where the lightest gray represents the lowest numerical group and solid black is the highest numerical group. The intermediate groups are represented in increasing numerical order by increasing darkness between the light gray and black groups. . An example of this type of display is in Figure 5. Figure 5 is the GVW calculation of Figure 4.

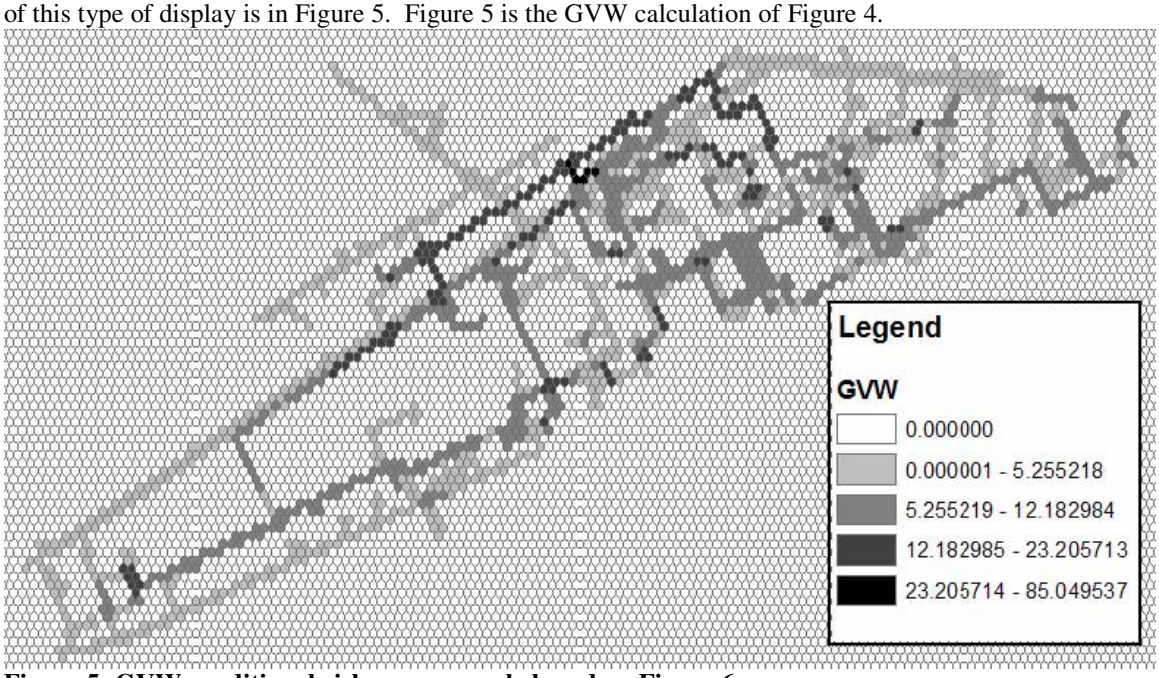


Figure 5: GVW conditional risk map example based on Figure 6

The color ramp is performed using Jenks Breaks of the GVW data and is a GIS internal algorithm that finds groups of numbers. This is *not* a linear ramp. It finds natural breaks in the data by finding groups of close numbers while maximizing the distance between these groups. This distance is dependent on the number of groups the user asks the program to create. With a given number of groups, the algorithm places the boundaries of the groups where there are relatively large jumps in the data, thereby maximizing the distance between the groups. For our purposes, it easily breaks out high value groups of GVW hexagons as well as other smaller groups.

Due to the nature of Jenks Breaks the groups depend on the values in the data. Therefore, if multiple Jenks Breaks are presented on a map for multiple data sets, the Jenks Break's groups created are

only valid for their respective data set. For example, in Figure 6 there are three different infrastructures presented simultaneously. In this figure, each of the three infrastructures has a Jenks Breaks preformed on its own dataset, therefore though not obvious, the figure has three color ramps displayed simultaneously, one for each infrastructure. The difficulty in observing these three different infrastructures at the same time gives a good basis for why we perform the GVW calculations and present them in their own map. This new map presents a spatial combination of all three infrastructures, e.g. Figure 5 for the case of Figure 4's infrastructures. We the reason the valued worths of the three infrastructures do not have the same number range is due to the network layout, number of users, and the disutility levels its unavailability creates to the users it supplies. Therefore the high and low areas of each infrastructure independently are important to the stakeholders of that particular infrastructure. Yet, when we apply a Jenks Break to the GVW in Figure 5 we see that, for example, the high group (black) locations are different than the high group locations of the individual infrastructures in Figure 4. Again, this is something not readily observable by looking at Figure 4 with its three different infrastructures. The GVW covers a different number set therefore the Jenks Breaks groupings are different.

This whole methodology can be computationally time consuming until after the Monte Carlo simulations have been run. However, at this point, we can easily change the size of the radius of influence to change the grid size and get new GVW within seconds. The new GVW is conditional on the new radius of influence.

We will now apply the time-independent methodology to the MIT campus.

4. MIT Case Study

We used the MIT campus as a case study due to the accessibility of infrastructure layouts. We had an open dialog with the MIT Department of Facilities (DOF), which provided us with information regarding specifics of the infrastructures. The case study was originally preformed on the real MIT infrastructure layouts. Interpretation of the results made intuitive sense and yielded consistent results with our value tree. These real infrastructure results are *not* presented here. For security purposes, we have changed the geographic layouts of the infrastructures. However, we kept several things the same to yield results that we could interpret. These are: the user PI; the number of arcs and nodes; the number of users serviced by an infrastructure; the number of supplies per infrastructures particularly the main on-campus generated resources.

We will use the same value tree as Apostolakis and Lemon (2005). This value tree considers an MIT building a user, thus also we consider any building under MIT jurisdiction a user. There were 133 building covered by this case study. Not all users were supplied by all infrastructures and some users were supplied redundantly. The constructed scales were developed by Apostolakis and Lemon. We have kept their zoned levels but modified the descriptions to fit our MG scheme.

The case study is presented in the following fashion. First, we present a background on the MIT infrastructures we have chosen to analyze. Then, we present the MG grouping, value tree, constructed scales, and weighting functions as well as their results (user PIs). Following this, we present the results of each infrastructure independently, i.e., their valued worth. We then present the GVW of the MIT campus.

4.1. MIT Infrastructure Background and Analysis Setup

We selected the following infrastructures for the case study: Chilled Water Supply, Domestic Water Supply, Steam Supply, Natural Gas, and Electric Power. We felt that telephone and data systems, which can also be wireless, presented a different set of problems that could not be addressed through the type of analysis presented here because their users and sources are not well defined. We also ignored the "return resources," i.e., chilled water return, sewer and storm drain, and steam condensate, again due to a more complicated user/supply scheme.

The physical networks for the analyzed infrastructures were taken directly from the DOF CAD drawings and are presented here with modifications for security purposes. Again, for this study, all of the campus buildings were considered users. The physical place where we placed the user nodes for each network on our GIS maps is where the MIT DOF campus wide plans reference to a building floor plan. This occurs, in the majority of cases, after the first isolation valve or access node within the physical building.

Several buildings, for various infrastructures, are serviced by multiple sources or multiple

pipes/lines from the same source. Given the same infrastructure, all of the service lines ending within a building are considered to service the same user, and a loss of a resource to that building would require the loss of that resource through all lines of the respective infrastructure servicing the building. Also, in some places, resources do not physically connect to a building but still service it; we assume the closest building to the point user to be serviced by the resource. For example a fire hydrant outside of buildings would be considered a fire resource for the nearest building.

For the most part, flow direction was not considered in this analysis but the coding has been done for this and will be implemented in later research/results. The only part of this research where there is directed flow is where it is labeled thusly on the CAD drawings, which only occurs in the steam system. Also, dead-end/end-capped resources are modeled to the point where they are isolated from the rest of the system by a valve. Non-user nodes are placed where there is any access to or equipment for the system. Examples are manholes, hand holes, meters, isolation valves, switch boards, etc. Nodes are also located at any intersections of piping.

4.2. Background notes on MIT infrastructures

Electric:

- Central Utilities Plant (CUP) can supply up to 80% of full power
- Extra power is purchased and supplied offsite
- There are two main buses which service four major loops and each bus services one half of each loop under normal conditions

Domestic Water:

-All domestic water is supplied from several pipes from off campus suppliers

-The domestic water supplies water to the fire suppression system and there are several dedicated fire suppression loops

Steam:

-MIT produces all of its steam from the Central Utilities Plant

Natural Gas:

-All natural gas is supplied from off site on several lines

Chilled Water:

-MIT produces all of its own chilled water

-The main supply is from the Central Utilities Plant

-There is a smaller plant that is located and services several building on east campus

4.3. MIT Value Tree and User PI assessment

There were 133 building analyzed and each building is considered a user. Not all buildings receive all infrastructure resources; major examples of this are the chilled water and steam systems. Due to the large number of users, we did not want to analyze each user individually. Instead, we setup a hierarchy of users. For MIT, we divided the buildings into four macro-user groups (MG): Residential (dorms, etc), Academic and Research (classrooms and laboratories), Support Facilities (utility plants, etc), and Athletics (sports, gym, etc). We later assigned different disutilities, via constructed scales, to the PMs of each of these MGs. All users take on the same constructed scale levels as their respective MG. Figure 6 shows a graphical representation of where the different MGs are located throughout the campus. Note where the MGs group up: Residential in the bottom-left, Academic in the middle to upper-right, Athletic in the middle, and Support littered throughout. The MGs are also summarized in Table 6.

MG Type	Number of users	Total number of people in MG/Average	Total Floor Space in MG/Average (ft^2)	Total Lab Space in MG/Average (ft^2)
A&R	68	9,800 / 144	3,506,898 / 51,572	1,253,550 /18,434
Residential	26	8,350 / 320	1,624,816 / 62,493	2,739 / 105
Support	29	710 / 24	846,328 / 29,183	30,710 / 1,059
Athletic	10	850 / 85	340,226 / 34,022	0 / 0

Table 6: Summary of Macro-User Groups

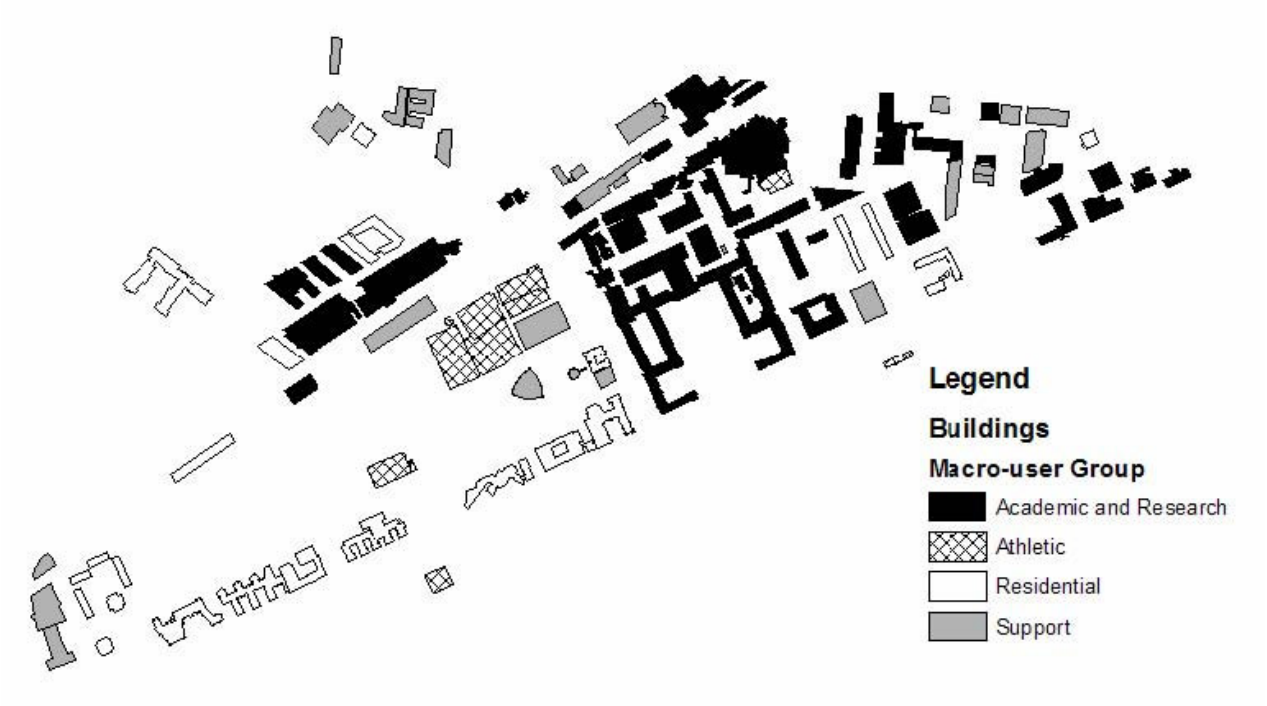


Table 7: Map of Users given by Macro-user Group

We then had to develop constructed scales in order for the decision maker to assess the disutility evoked by events to users. These constructed scales were developed by the DOF at the same time the original value tree was created (Karydas and Gifun, 2002) and were modified by Apostolakis and Lemon (2005). To keep our answers consistent with theirs, we use the same leveling scheme but redefine some of the definitions so they are applicable to all of our MGs. For instance, in our constructed scale for impact of people (Table 5) we have added items like job loss and employment interruption to areas of Apostolakis's and Lemon's constructed scale for impact on people (Table 1). We essentially broadened the scope of impact on people particularly because A&R as well as support buildings would have been left without a way to assess their employment function. Our redefined constructed scales are presented in Table 5 and Table 8 through Table 14.

Level	Description	Unweighted Disutility
3	Major Environmental Impact	1.00
2	Moderate Environmental Impact	0.34
1	Minor Environmental Impact	0.04
0	No Environmental Impact	0.00

Table 8: Constructed Scale for environmental impact

Level	Description	Unweighted Disutility
3	Catastrophic physical property damage	1.00
2	Major physical property damage	0.27
1	Minor physical property damage	0.03
0	No physical property damage	0.00

Table 9: Constructed Scale for physical property damage

Level	Description	Unweighted Disutility
4	Extreme Interruption; Greater than 6 months, entire buildings evacuated and activities relocated.	1.00
3	Major Interruption; 1 to 6 months, laboratories evacuated and activities relocated.	0.57
2	Moderate Interruption; 1 to 4 weeks, specialty classrooms evacuated and activities relocated.	0.19
1	Minor Interruption; Less than 1 week, a few administrative units or small classrooms evacuated and activities relocated.	0.06
0	No Interruption	0.00

Table 10: Constructed Scale for interruption of academic activities & operations

Level	Description	Unweighted Disutility
3	Catastrophic intellectual property damage; Long-term experiments	1.00
2	Major intellectual property damage; Artifacts and rare documents	0.46
1	Minor intellectual property damage; Non-backed up electronic data	0.05
0	No intellectual property damage	0.00

Table 11: Constructed Scale for intellectual property damage

Level	Description	Unweighted Disutility
3	Major degree of adverse publicity; Petitions, demonstrations	1.00
2	Moderate degree of adverse publicity; Negative articles published	0.34
1	Minor degree of adverse publicity; Verbal complaints	0.04
0	No adverse publicity	0.00

Table 12: Constructed Scale for internal public image

Level	Description	Unweighted Disutility
3	Major degree of adverse publicity; Affects enrollment, contributions, program funding, or faculty recruiting	1.00
2	Moderate degree of adverse publicity; National/International Media	0.57
1	Minor degree of adverse publicity; Local media	0.06
0	No adverse publicity	0.00

Table 13: Constructed Scale for external public image

Level	Description	Unweighted Disutility
4	Extreme Impact on projects, funding, employment, and students	1.00
3	Major Impact on projects, funding, employment, and students	0.50
2	Moderate Impact on projects, funding, and students	0.23
1	Minor Impact on students	0.02
0	No Impact	0.00

Table 14: Constructed Scale for programs affected

As we discussed in Section 3.1, to diversify the users even more we added what we called the weighting function. This involved multiplying a natural scale by a constructed scale. Adding this natural scale which is continuous and based on physical data about the users adds another level of distinction between users. These data are attributes already assigned in the GIS database to the buildings. They include the maximum human populations of the buildings, laboratory area, and total usable area of the building. Classroom area was assumed to be usable area minus lab area in academic buildings. These data created a scaling effect on the constructed scales. With these data we added weighting functions to all of the PMs we analyzed with the relation in Table 15.

Data from GIS	PM for weighting function
Max occupancy	People, Programs Affected

Lab square footage	Environment Damage, Intellectual Property, Programs Affected
Classroom square footage	Interruption of Academic Activities
Usable square footage	Physical Property Damage

Table 15: Data and respective PM for weighting function use at MIT

All of our functions were linear and can be seen in column 2 of Table 16 for each PM. For example, for the PM physical property damage, our weighting function was usable square footage of the user divided by the maximum square footage user. Our largest usable area was 286,527 ft^2 and an example building A (an A&R building) had a usable area of 51,979 ft^2, therefore building A's weighting function value was 0.1814. This weighting function was multiplied by the picked zoned level for the physical property damage PM for an event. For example, the A&R loss of a steam resource is expected to cause minor damage to the building (level 1) which corresponds to a disutility d = 0.03 (Table 9). Therefore, the disutility for the physical property damage PM for building A was 0.03*0.1814 = 0.0054. Once the disutilities for the other PMs for building A are calculated for the loss of the steam system, the values are added together via Equation (1) to create the total PI for user A for the event of loss of the steam system.

The actual PIs for each user were assessed for each infrastructure loss according to Eq. (1). There are several buildings on campus, e.g., the Research Reactor, that would normally warrant special attention to assess its value or even have its own analysis preformed on it. For our purposes, we try to classify these building into our scheme using the same MG groups, but recognize their importance.

We assessed the constructed scales of each PM for each MG for the loss of each of the five infrastructures independently. We took several samples of user PIs for various infrastructure losses to compare and consistency check them. For example, in Table 16 we present an A&R building for the loss of gas compared to an Athletic building for the loss of steam. Since we have included weighting functions the users must be checked for consistency in the context of their user data (population, lab area, useable area). Therefore, if we look at the final PI, we see that they are very close, thus suggesting that we should be almost indifferent between the two different events (one user for gas loss and the other for steam loss) to the respective users in the context of not only their MG, but also the size of the building, people in the building, and lab space in the building. We also showed these consistency checks to an individual whom was present at the original MIT DOF stakeholder deliberations.

Figure 6 compares all users grouped by the 4 MG PIs for all five events (loss of each infrastructure) (note the PI is cumulative for all events). It is notable in Figure 6 how the MG hierarchy via picking different constructed scales created diversity between each MG based for a given event. The weighting function then diversified each user within an MG based on the GIS data.

User Data	
User *a*	**User *b***
MG: Athletic Event: Steam Loss Population: 42 Lab Area (ft^2): 0 Useable Area (ft^2): 16613	MG: A&R Event: Steam Loss Population: 220 Lab Area (ft^2): 57685 Useable Area (ft^2): 66069

PMs (weight (w))	Weighting Function for PM	Disutility Calculations	User	
			User *a*	**User *b***
Impact on People (0.295)	$\frac{User\_Population}{1200}$	Zoned Level	1	1
		Unweighted Disutility	0.05	0.05
		Weighting Function Value	0.03461	0.18352
		Weighted Disutility (d)	**0.00173**	**0.00918**
Impact on Environment (0.196)	$\frac{Lab\_Area}{119012}$	Zoned Level	0	1
		Unweighted Disutility	0	0.04
		Weighting Function Value	0	0.835869
		Weighted Disutility (d)	**0**	**.0334**
Physical Property Damage (0.049)	$\frac{Useable\_Area}{286726}$	Zoned Level	0	1
		Unweighted Disutility	0	0.03
		Weighting Function Value	0	0.230586
		Weighted Disutility (d)	**0**	**.00692**
Impact on Academic Programs and Operations (0.056)	$1-\frac{Lab\_Area}{119012}$	Zoned Level	0	1
		Unweighted Disutility	0	0.06
		Weighting Function Value	0	0.1461
		Weighted Disutility (d)	**0**	**.008766**
Intellectual Property Damage (0.128)	$\frac{Lab\_Area}{Useable\_Area}$	Zoned Level	0	1
		Unweighted Disutility	0	0.05
		Weighting Function Value	0	0.873102
		Weighted Disutility (d)	**0**	**.04366**

Impact on External Image (0.083)	N/A	Zoned Level	1	1
		Unweighted Disutility	0.06	0.06
		Weighting Function Value	N/A	N/A
		Weighted Disutility (d)	**0.06**	**0.06**
Impact on Internal Image (0.055)	N/A	Zoned Level	2	1
		Unweighted Disutility	0.34	0.04
		Weighting Function Value	N/A	N/A
		Weighted Disutility (d)	**0.34**	**0.04**
Programs Affected (0.138)	$\frac{\left(\frac{Lab\_Area}{119012}+\frac{Population}{1200}\right)}{2}$	Zoned Level	1	1
		Unweighted Disutility	0.02	0.02
		Weighting Function Value	0.017305	0.509697
		Weighted Disutility (d)	0.000346	0.01019
USER PI			**0.024238**	**0.024325**

Table 16: Example of a user consistency check

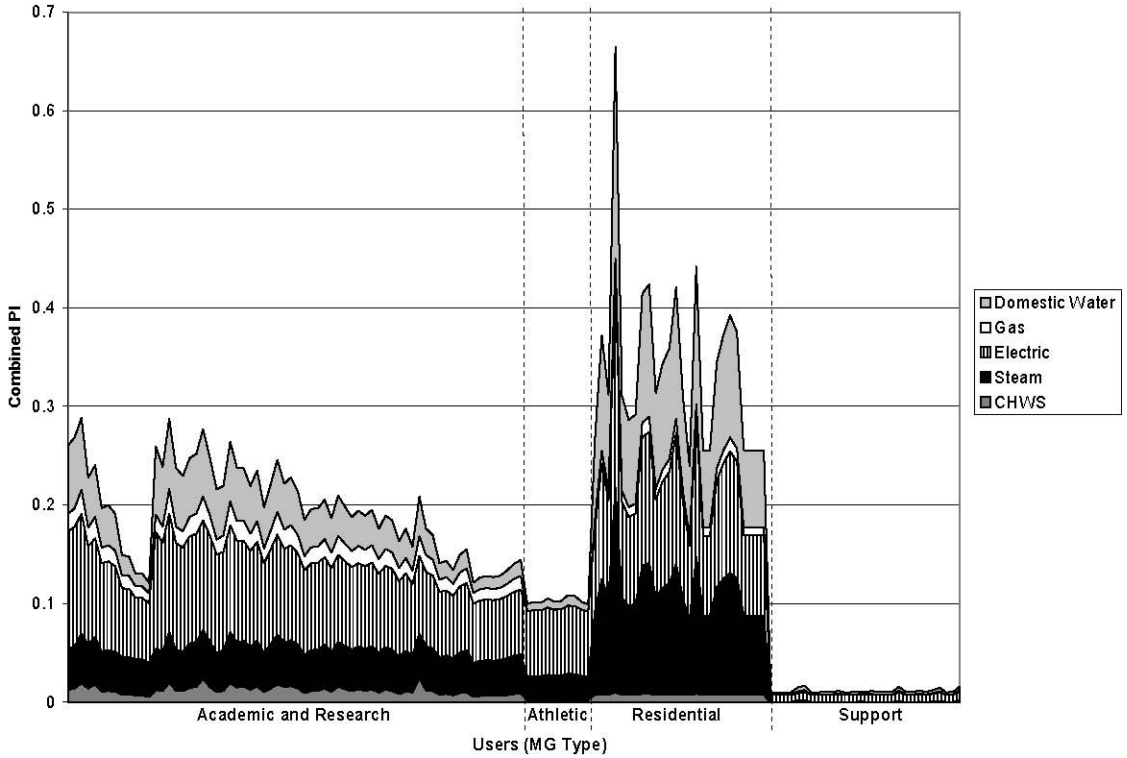


Figure 6: All users grouped by MG for comparison for all events. The PI on the y-axis is cumulative; each event's PI is added to the others. To get an individual event's PI for a user take look at only the height portion the respective shaded area covers.

The following are some major trends in the resulting PIs: 1) All A&R users have high PI for electric power loss; 2) the more classroom percent space in an A&R user, the closer the PI for electric power loss and steam loss get, i.e., heating is relatively more important to classrooms than labs; 3) Residential MG users' highest PI vary among domestic water, steam, and electric loss; 4) Resident users are relatively close in their PI for domestic water, steam, and electric power loss, resident buildings with computer labs have a high PI for electric power loss; 5) Support buildings have the highest PI for electric power loss; 6) Athletic buildings have the highest PI for electric power loss; 7) The large spikes in the residential MGs are caused by the most populated building on campus; 8) Support buildings do not evoke much disutility for all events mainly due to the small size of the buildings and the low population of people in them; 9) Athletic users are the second lowest group due to the inability to cause disutility for events to many of the PMs, e.g., academic operations, intellectual property, etc..

The calculated PIs just developed for each of the five events (loss of electric, domestic water, natural gas, steam, and chilled water) were inputted into the GIS database for each user; i.e. each user has five PIs, one for each event. We then performed the network analysis on each infrastructure.

4.4. Network Analysis

A Monte Carlo network analysis was performed in accordance with the methodology covered previously. The first step in simulation is to first pick our probability of a successful attack. We chose a low number (0.01). With this number set, we had to simulate enough trials to make our results reliable. To make what we considered a reliable simulation, we took relative variance of a random test element for each users simulations (Billington, 1992). We chose a 5% relative variance† as our goal. The test element was chosen randomly for each infrastructure and was based on the RAW importance measure value. Simulations of 20,000 trials were not consistently under 0.05 relative variance. However when we ran 30,000 random system states for each user where all elements had a 1% chance of a successful attack, we got a relative variance under 5% for the test element. Thus 30,000 simulations with a 1% change of a successful attack were our Monte Carlo simulation parameters. With the parameters set we began the simulations.

The program chose random numbers to simulate the 30,000 states for each user. We tracked the U_{total}, $U_{system\text{-}fail}$, $U_{k\text{-}fail,}$ $U_{k\text{-}fail\text{-}system\text{-}fail,}$ $U_{k\text{-}working,}$ $U_{k\text{-}working\text{-}system\text{-}fail}$ counters until the user/infrastructure combination and created the worth for each element according to Equation (10). Once all users for one infrastructure were simulated we calculated the valued worth for all of the current infrastructure's elements according to Equation (11).

4.5. Individual Infrastructure Results

Before we move on to show the GVW, we will look at the resultant VW's of each infrastructure individually.

4.5.1. Chilled Water System

The VW shown in Figure 7 is for the chilled water system (CHWS) using RAW.

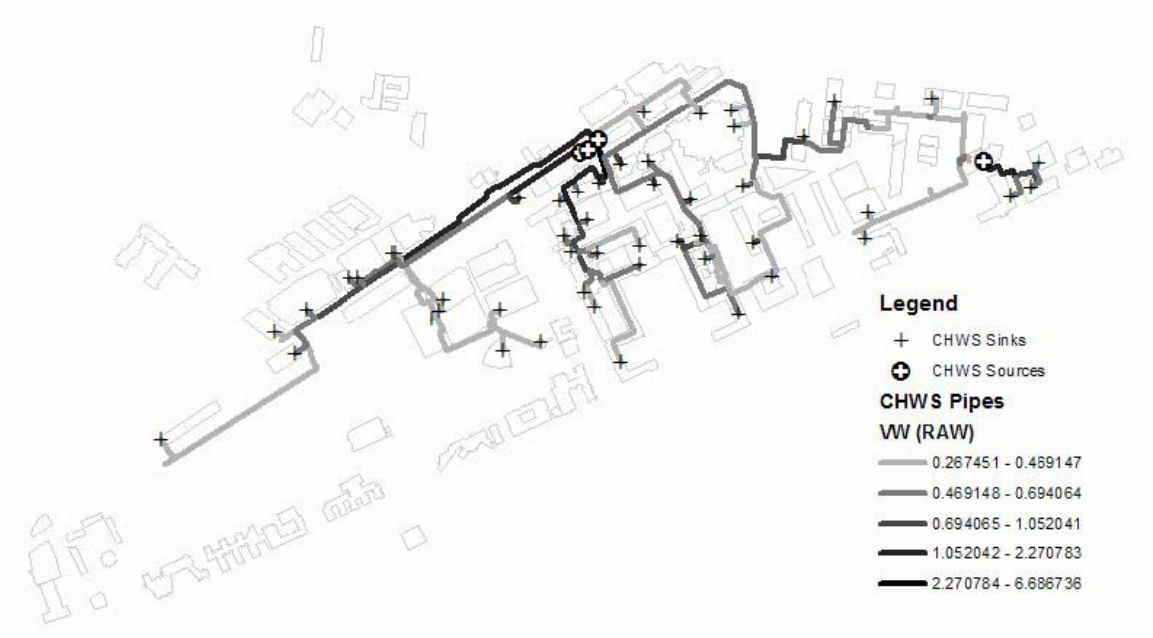


Figure 7: Chilled Water System VW using RAW

Looking at Figure 7's network layout and where the supplies and users are located, we can check our results. The highest VW areas are pipes entering the Central Utility Plant, which, aside from a minor plant to the right, is the only production source of chilled water on campus. The higher areas in general surround the academic buildings, if we look at Figure 6 we see that A&R users have the highest PI for CHWS loss and thus created higher VW. There are some redundant loops which lower the VW in the middle of the campus. However for the most part this system is mostly in series and filters through very few pipes coming out of the CUP.

The number set of the VWs is 0.267451 – 6.686736 which is relatively low when compared to the other infrastructures. Despite the system's lack of redundancy, the loss of CHWS to its users does not cause much disutility to them. Thus we would expect a relatively low number set.

4.5.2 Electric Power System

The VW shown in Figure 8 is for the electric power system using RAW.

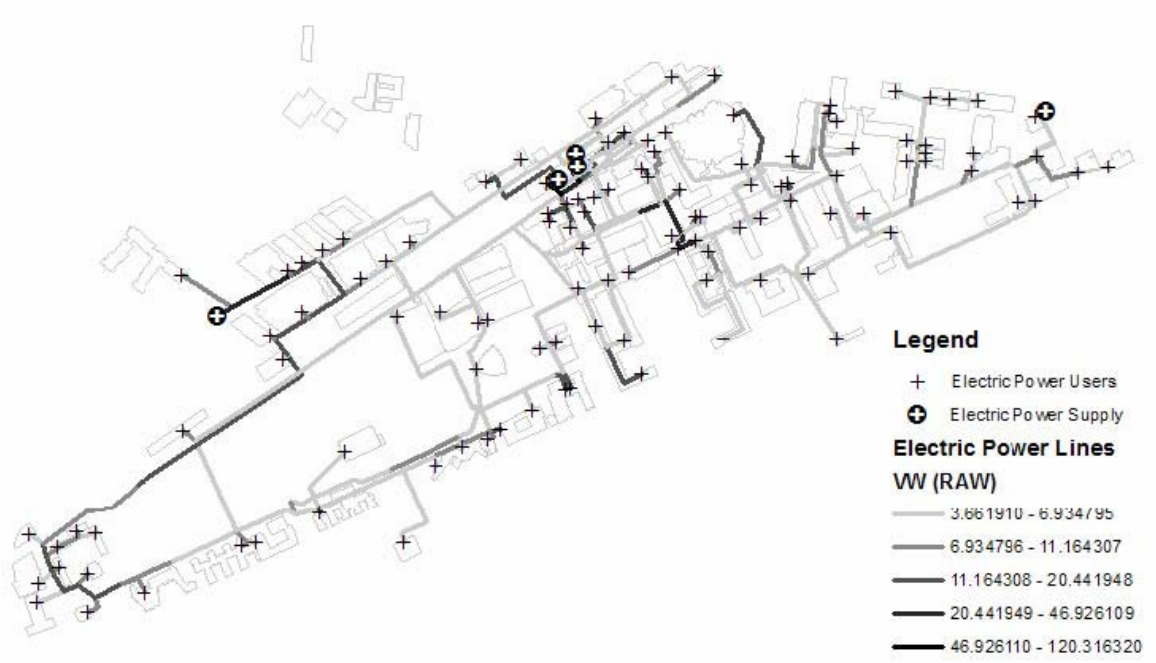


Figure 8: Electric Power System VW using RAW

We check our results again. We see that our highest valued lines are coming out of the CUP or are located toward the left side of campus. The left side supplies the residential users and is therefore more important than some of the other areas. However, we really need to consider the VW number range of the electric power system. The number set of the valued worths is 3.661910 – 120.316320, which contains the highest values when compared to the other infrastructures. All users rely heavily on the electric power system. This system, despite its four separate loops, is not very redundant along its loops.

The number range suggests that losing our least valued electric power line is almost as important to us as our top natural gas lines (see next section). It also shows that the majority of the lines are more important than most all of the CHWS pipes, all of the natural gas pipes, most of the steam pipes, and most of the DOMW pipes. Thus, despite the relative ranking of the electric power lines being higher or lower than one another, when the GVW is found in the next section, these "low ranked" lines will prove to create high GVW areas for the whole campus. Therefore, the results for the electric power system should prove to dictate much of the GVW.

4.5.3 Natural Gas System

The VW shown in Figure 9 is for the natural gas system using RAW.

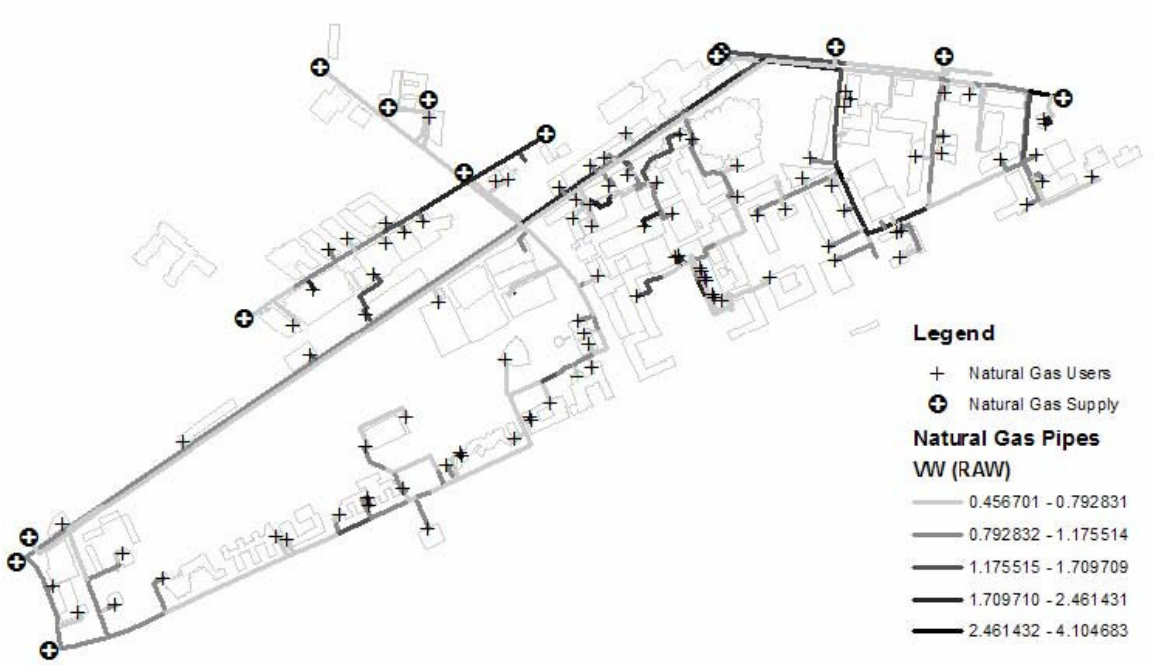


Figure 9: Natural Gas System VW using RAW

We check our results again. This system has the most supplies and therefore a lot of redundancy. This causes a lack of a concentration of high VW areas. Individually, the A&R and residential users tend to rely on natural gas equally. However, there are over twice the number of A&R uses as residential so the higher valued areas tend to center around the concentration of A&R users in the main campus (center to right of the map).

The number rank of the valued worths is 0.456701 – 4.104683, which is relatively low when compared to the other infrastructures. Again, looking at Figure 6, we note that there is not much disutility caused by the natural gas so we would expect lower values. The values are even lower than the CHWS due to the CHWS centralized supply area.

4.5.4 Domestic Water System

The VW shown in Figure 10 is for the domestic water system (DOMW) using RAW.

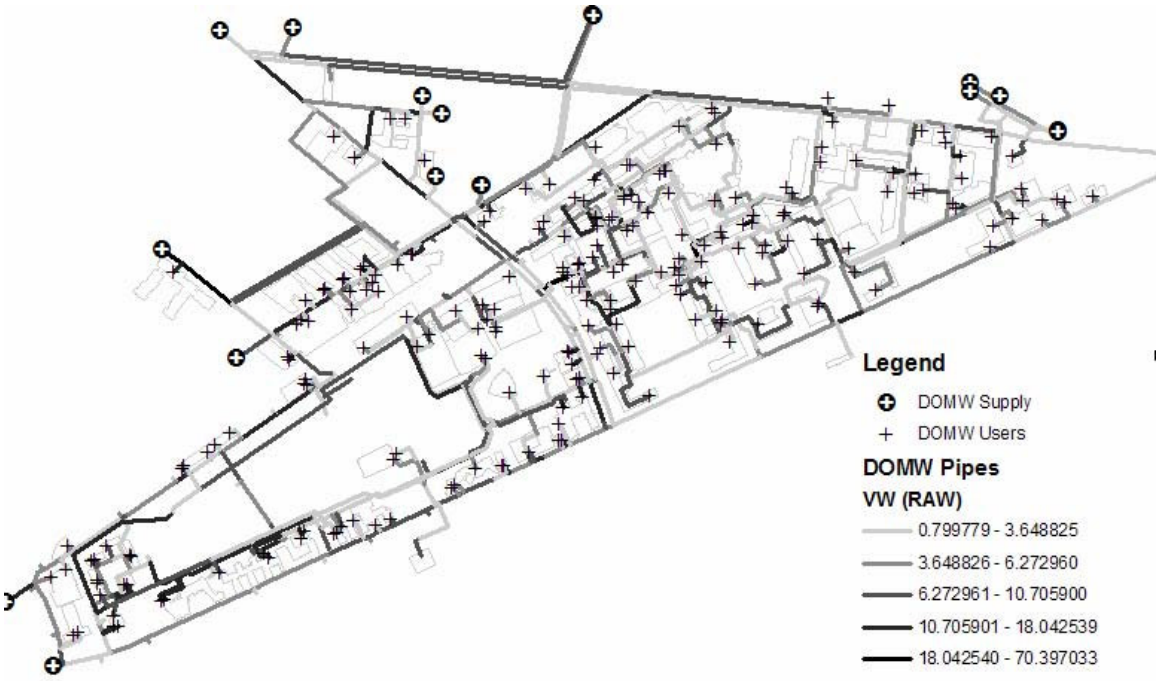


Figure 10: Domestic Water System VW using RAW

Like the natural gas infrastructure, we have many different supplies from off campus feeding our users. The large mains have redundant connections to each other. However, once the network draws away from the main lines, there are many series connections. Yet, since these connections do not serve as many users as the main supply pipes, we find that the highest areas are the pipes near the supplies.

The large dependence of residential users on DOMW creates high VW in the lower left of our map. We can also note that the dinning areas (residential MG) on campus have high VW pipes servicing them due to their reliance on DOMW to feed the campus.

There is also a concentration in the main A&R buildings (dead center of the map) which have numerous classrooms and therefore numerous people are located in these areas. Thus, they have a high reliance on DOMW particularly for the bathrooms.

The number set of the valued worths is 0.299779 – 70.397033, which is relatively high when compared to the other infrastructures. Again, looking at Figure 6 we note that there is a lot disutility caused particularly by the residential users. The extra redundancy particularly in the middle of the campus creates VW numbers that are on par with some of the CHWS and natural gas pipes. This is the most redundant system.

4.5.5 Steam System

The VW shown in Figure 11 is for the steam system using RAW.

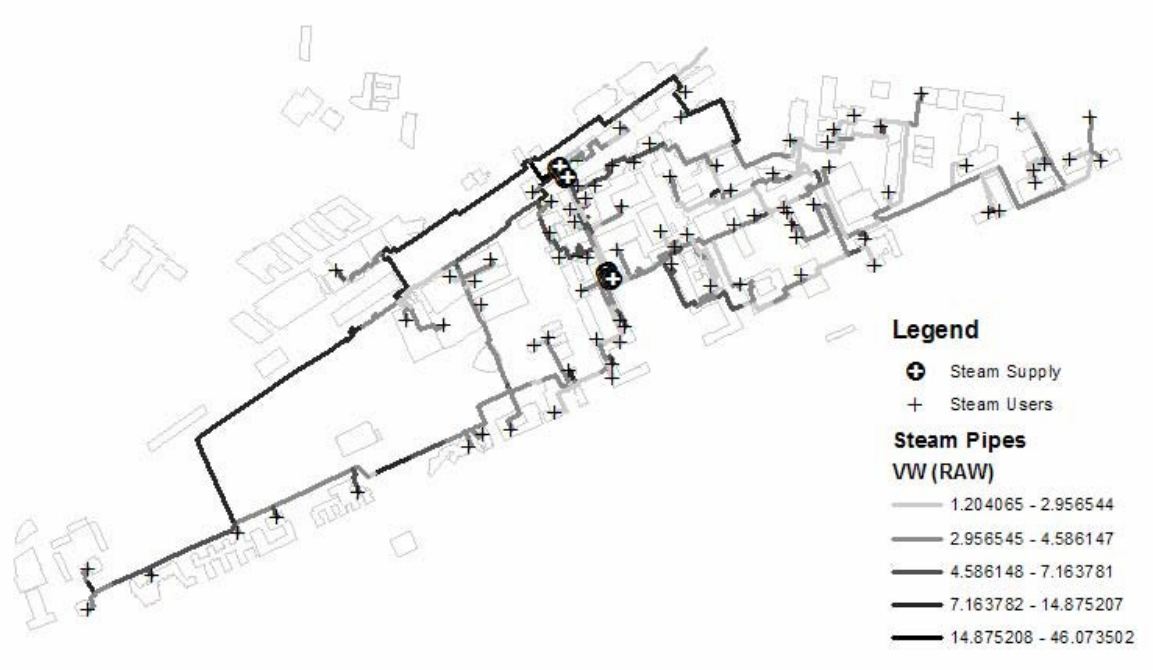


Figure 11: Steam System VW using RAW

Like the CHWS, the steam system is an on-campus utility. Therefore, we find a concentration of high VW pipes around the Central Utilitys Plant. But the loss of steam can cause high disutility to the residential users on the left side of campus so much of the high VW pipes are locate more to the left side. The left side of the campus is also not very redundant.

We can see that the upper supply out of the CUP is higher VW than the lower since the lower and eastern users on campus are more redundantly supplied by a second steam plant just below the CUP, almost in the center of the map.

The number set of the valued worths is 1.1204065 – 46.073502, which is the middle infrastructure as far as the VW are concerned. But 46 is a high VW which is due to the large dependence of the residential users and some A&R users on steam while also being less redundant to the residential users. We can see that the lowest values of the steam system are not as low as, for example the DOMW system, due to the lack of redundancy in the system.

4.6 Geographic Valued Worth

As explained in the methodology section, we must pick a threat that our GVW can be conditional upon. We have chosen a bomb that can destroy everything in a 7-meter radius. This 7-meter radius is our radius of influence and all GVW results will be conditional on it. To develop the GVW a grid of hexagons with the height and width of two times our radius of influence was developed and the GVW was calculated. This result is in Figure 12.

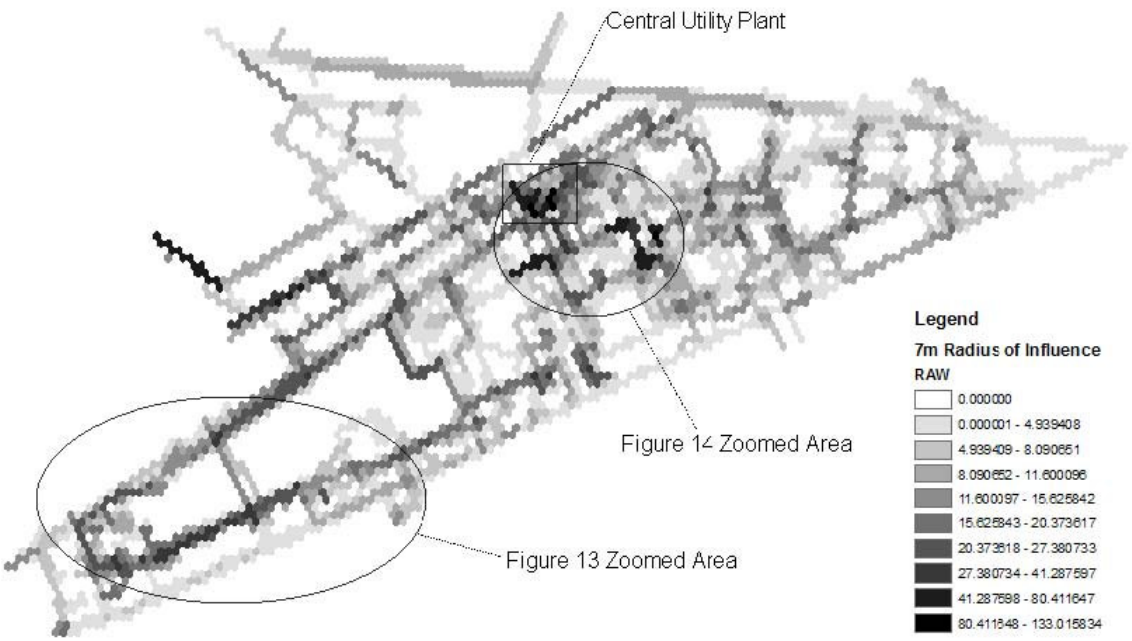


Figure 12: GVW Conditional Risk Map using RAW for the MIT campus

The highest GVWs are located near the Central Utility Plant, which makes sense since the CUP produces 80% of the electricity, all of the steam, and all of the chilled water for the campus. This means the immediate areas just outside of the plant are of great importance to the campus. We note that the high GVW values follow much of the electric power system, which makes sense since it causes the highest disutility to most users.

There is a high GVW "loop" that services the residential users in the bottom left of the campus. If we zoom in (Figure 13) on this area, we see that there are pipes for electric power, steam, water, and natural gas that all run under the same streets very close to each other, many times within the radius of influence, till they service the dorms in the left campus. These long stretches are not very redundant.

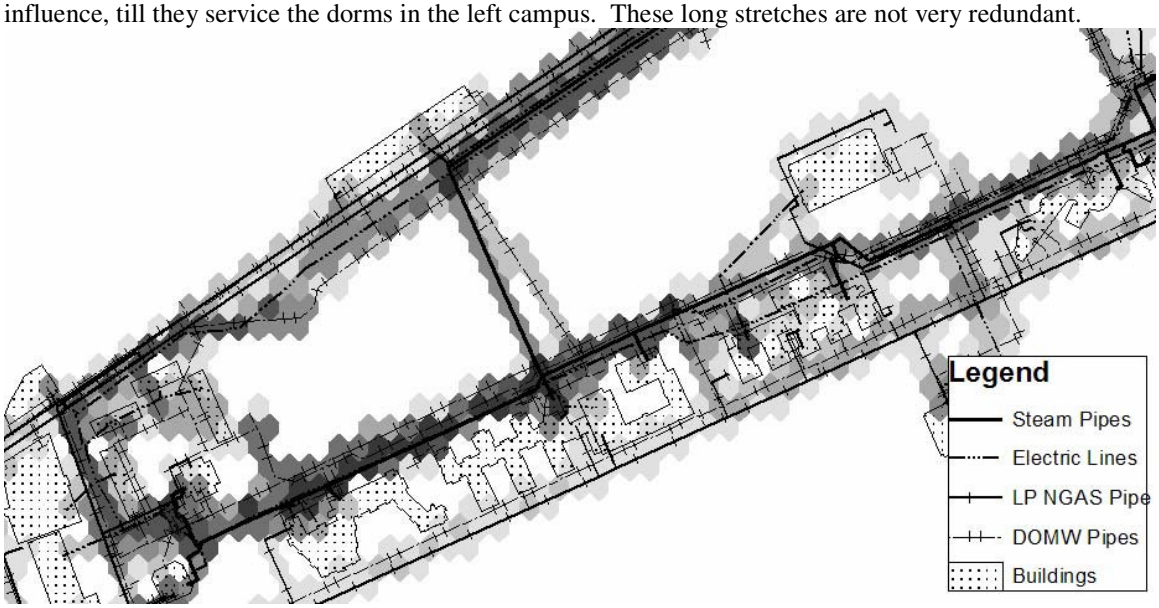


Figure 13: Zoom in on left size of campus. GVW in the background.

We also note the high GVW area in the left side and middle of campus. This is caused by the high

VW of the DOMW system, which services a major dorm and a few major A&R buildings without much redundancy.

Also in the middle of campus we see two areas below and to the left and right of the CUP (Figure 14). All five infrastructures pass though these areas to service the left and right sides of campus. Therefore, these two areas are two low-redundancy bottle necks, or choke points, on the campus.

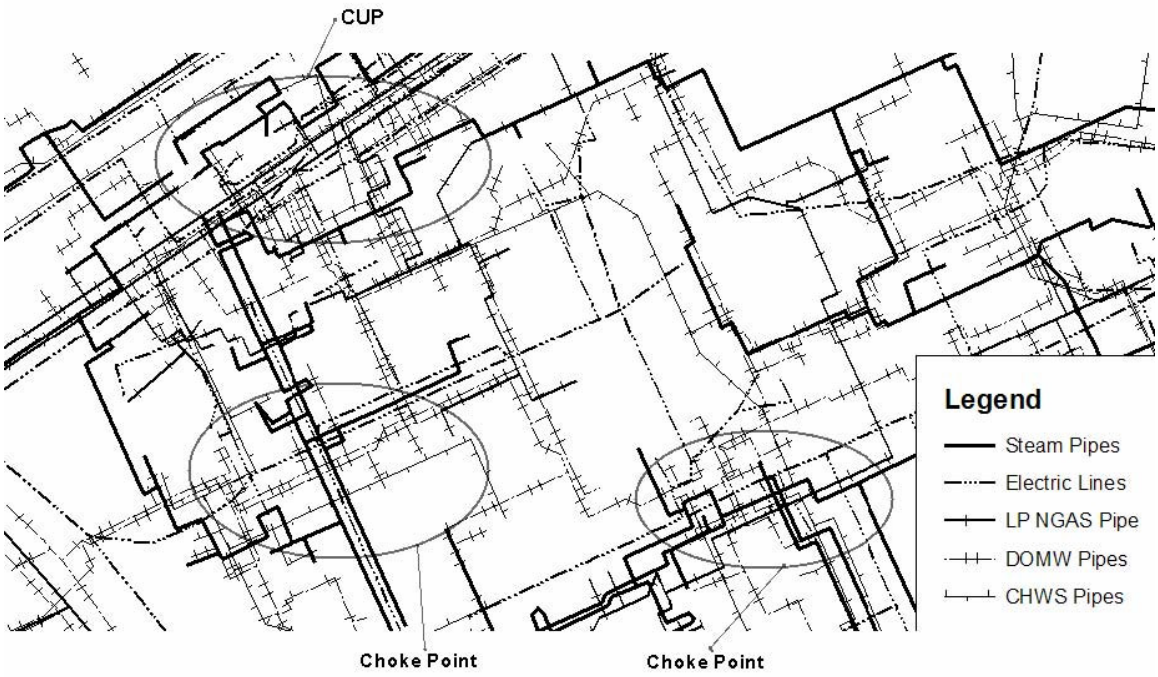


Figure 14: Zoom in around CUP and two choke points in center of campus. No GVW displayed.

We see that the proximity of infrastructures within a radius of influence leads to larger GVW values, as it should. The proximity of the electric system to the steam or domestic water system in the left campus creates the highest GVW areas. The proximity of steam and DOMW pipes in the left campus can create areas more risky than the highest electric power areas when considered independently. Gas and CHWS do not affect the GVW too much since their additions do not raise the GVW high enough to compete with the high values where the DOMW, electric, or steam systems are collocated. However, they can influence middle ranked pipes of the DOMW, electric, or steam to move the area up in GVW rank enough to become a notable GVW area.

We must remember that by using importance measures the solution we are giving is not like that in a minimal-cut-set analysis. We cannot say that the highest valued worth (geographic or infrastructure element) will for sure cause a failure to its connected users. Rather, we determine that the higher the GVW the more the unavailability of the elements passing through it the closer the systems are to failure with a certain amount of disutility. The elements in the area may or may not cause directly a resource loss to the respective users.

5. Input to Decision Makers

The conditional risk maps are given to decision makers to decide the susceptibility of the areas to attack. Since all of our infrastructures have the same failure probability, the decision makers must make these susceptibility decisions. In time, we could apply susceptibility before the GVW analysis by changing the probabilities of failure. We could for instance add a program which would determine the probability of failure of an element conditional on the threat, e.g., the amount of TNT a bomb has, and how the element is located, e.g., buried under concrete vs. under soil. GIS has the capability of doing 3D analysis, therefore we could affect underground lines and above-ground lines of varying heights and depths differently. If an area were determined to be of high GVW and high susceptibility by the decision makers, then it should be considered a critical location of the entire system, i.e., for all infrastructures and users. The GVW is most

important to the decision makers because it is a global metric that represents the worth (to the decision makers) of a location across all infrastructures. The GVW does not care about the infrastructures individually, only the result of combining all infrastructures. In a single-infrastructure analysis, we find elements of significant risk for that particular infrastructure, like our VW results. If we were allocating resources to protect that infrastructure only, we would apply the resources to those high-risk elements. However, in a terrorist scenario we are trying to protect society and what society (i.e., the decision makers) deems important. Therefore, our multi- infrastructure analysis and the GVWs prioritize areas that are important to society so that resources can be allocated accordingly.

The highest risk areas of two infrastructures, when analyzed separately, are in general different from those that are found when these infrastructures are analyzed together, as we did in this paper.

.

6. Conclusions

The methodology we have presented takes the results of stakeholder deliberations about the performance measures that the stakeholders deem important to society and converts their values through network and spatial analysis to a ranking of geographic areas that can adversely disturb the infrastructure services to the stakeholders. It accomplishes this by first determining a valued worth of each of the elements of the infrastructures networks. This valued worth is based on the characteristics of the users (PIs) and the importance measures of the elements which make up the infrastructures which supply resources to those users. The methodology then assesses the geographic valued worth of physical areas of defined size (based on the radius of influence) by determining combining the values of all the infrastructure elements within the physical area. The result is a ranking of physical areas that can be expressed in a graphical form on a map. This map aids decision makers in the allocation of countermeasures to better protect society from malicious threats.

We note that we have made some broad assumptions (user independence, PM independence, failure probabilities) in several areas, however, with expert opinion the assumptions could be limited to provide a realistic analysis of the infrastructures. Despite the assumptions in our case study, we believe this method of developing GVW across multiple infrastructures for terrorism is of importance to decision makers for anti-terrorism resource allocation. The process we have presented effectively draws attention to the geographic areas that merit attention by the decision makers. Had we only relied on the analysis of the infrastructures individually, we might not have noticed that a geographic location, which may be of moderate importance to the networks independently, is of extremely high importance if all infrastructures are considered.

References

Apostolakis, G.E., and Lemon, D.M., 2005. "A Screening Methodology for the Identification and Ranking of Infrastructure Vulnerabilities due to Terrorism," *Risk Analysis*, 25:361-376.

Billinton, R., 1992. *Reliability Evaluation of Engineering Systems: Concepts and Techniques.* New York : Plenum Press.

Cheok, M.C., Parry, G.W., and Sherry, R.R..,1997. "Use of importance measures in risk-informed regulatory applications" *Reliability Engineering and System Safety,* 60: 213-226.

Clemen, R. T., 1996. *Making Hard Decisions: An Introduction to Decision Analysis.* 2nd Edition. Belmont, CA, Duxbury Press.

Garrick, B.J., J.E. Hall, M.Kilger, J.C.McDonald, J.C. McGroddy, T.O'Toole, P.S. Probost, E. Rindskopf Parker, R. Rosenthal, A.W. Trivelpiece, L. Van Arsdale, and E. Zebroski, 2004. "Confronting the Risks of Terrorism: Making the Right Decisions," *Reliability Engineering and System Safety,* 86:129-176.

Gregory, R., and Keeney, R. L., 1995. "Creating policy alternatives using stakeholder values." *Management Science*, 40:1035-1048.

Haimes, Y.Y., Horowitz, B.M., 2004. "Modeling interdependent infrastructures for sustainable counterterrorism." *Journal of Infrastructure Systems*, 10, 33-42.

Karydas, D.M., and Gifun, J.F. 2002. "A Methodology for the Efficient Prioritization of Infrastructure Renewal Projects." *Proceedings of the 6th International Conference on Probabilistic Safety Assessment and Management (PSAM 6),* San Juan, Puerto Rico, 23-28 June 2002, Editor: E.J. Bonano, Elsevier Science Ltd., United Kingdom.

Marseguerra, M. and E. Zio, 2002. *Basics of Monte Carlo Method with Application to System Reliability*.

Hagen, Germany: LiLoLe-Verlag.

Michaud, D., and Apostolakis, G.E., 2005. "Screening Vulnerabilities in Water-Supply Networks." In preparation.

National Research Council, 2002. *Making the Nation Safer,* National Academy Press, Washington, DC.

Office of Homeland Security, 2002. *National Strategy for Homeland Security,* U.S. Executive Office of the President, Washington, DC.

Paté-Cornell, M.E. and Guikema, S., 2002. "Probabilistic Modeling of Terrorist Threats: A Systems Analysis Approach to Setting Priorities among Countermeasures." *Military Operations Research,* 7:5-20.

Weil, R. and Apostolakis, G.E., 2001. "A methodology for the prioritization of operating experience in nuclear power plants." *Reliability Engineering and System Safety*, 74:23-42.

Zio, E., Podofillini, L., and Zille, V., 2006. "A Combination of Monte Carlo Simulation and Cellular Automata for Computing the Availability of Complex Network Systems," to appear in *Reliability Engineering and System Safety*.

.

Passwords, passwords everywhere, and not a minute to think!

Karen V. Renaud

Department of Computing Science,
University of Glasgow,
Glasgow, G12 9QQ.

Karen@dcs.gla.ac.uk
http://www.dcs.gla.ac.uk/~karen

We are being asked to remember increasing numbers of passwords and PINs, and for many computer users this load is becoming arduous if not unbearable. I will introduce some alternatives to passwords which do not rely on fallible recall memory and which place a much lighter load on the overburdened user.

I'll demonstrate some of the systems I have experimented with and discuss the pro's and con's of various alternative authentication mechanisms.

``Prospects for a Robust Electronic Voting Scheme for the UK''

Ishbel Duncan and Tim Storer,

Center for Digital Privacy, Security and Trust,
Department of Computing Science,
University of St. Andrews,
St. Andrews,
Scotland.

{ishbel,tws}@dcs.st-and.ac.uk

Using Computer Simulations to Support A Risk-Based Approach For Hospital Evacuation

C.W. Johnson,

Glasgow Accident Analysis Group,
Department of Computing Science, University of Glasgow, Glasgow, United Kingdom, G12 8QQ.
E-mail:johnson@dcs.glasgow.ac.uk http://www.dcs.gla.ac.uk/~johnson

Terrorist actions, such as the attacks on the London Underground and the Madrid train bombings, as well as fires, such as the destruction of the Station Night Club in Rhode Island, have focussed public attention on the evacuation of public buildings. Partly in consequence, there have been a number of recent legislative changes across Europe and the United States. This legislation encourages a risk-based approach to evacuation. Existing risk assessment techniques, including FMECA and fault trees, provide means of reasoning about potential fire hazards. They can also be extended to analyse the risks that occupants may not escape from a damaged building. However, it can be difficult to validate the findings from such analyses because group and individual behaviours have a profound impact on egress times. For instance, it is hard to assess the likelihood and consequence of the flocking behaviours that occur during mass evacuations. Live exercises address these limitations by providing direct insights into the behaviours of building occupants. However, these drills seldom recreate the conditions that hold during real emergencies, especially when occupants know that they are participating in an exercise. Ethical problems also restrict these drills. For example, patients' health can be jeopardised if they are evacuated from centres of care in a hospital. It can also be difficult to hold drills that might disrupt the 24/7 activities of power distribution and financial service companies. This paper, therefore, describes the development of the Glasgow-Hospital Evacuation Simulator (G-HES). G-HES is an interactive, stochastic software system that can be used to simulate the evacuation of large public buildings. It supports a 'risk-based' approach to evacuation and can be calibrated using observations from 'live' evacuation exercises. Managers can use it to explore the consequences of different staffing levels and evacuation procedures. Monte Carlo techniques provide means of calculating mean and worst-case evacuation times under these different conditions. The evacuation of a local general hospital is used as a case study. This decision is justified by the difficulty of performing such evacuations and by the relatively high number of fires that occur in hospital buildings each year[9].

Keywords: accident analysis; evacuation; simulation; human factors.

1. Introduction

Recent terrorist actions, such as the bombing of the London Underground, and the plethora of false alarms that follow such attacks have focused public attention on the evacuation of public buildings. Fires, such as the destruction of the Station Night Club in Rhode Island, have also increased concern. Partly in consequence, there have been considerable changes in the legal and regulatory frameworks that protect building occupants.

1.1 Regulatory Background

The United States provides both local and Federal regulations governing the evacuation of public buildings. Most states have adopted the provisions of the International Building Code, which requires that building records and floor plans show the "construction, size and character of all portions of the means of egress" (NCSBCS 2000, Section 106.1.2). The US Occupational Safety and Health Administration require employers to 'ensure that routes leading to the exits, as well as the areas beyond the exits, are accessible and free from materials or items that would impede individuals from easily and effectively evacuating'

[9] Thanks are due to F. Ashraf, J. Johnston, C. McAdam, G. Mckinlay and M. Wilson who drove the design and implementation of the simulation software described in the later sections of this paper.

(OSHA, 2003). The Code of Federal Regulations, Standard 29, Part 1910, Subpart E requires that employers prepare emergency action plans to address 'fire; toxic chemical releases; hurricanes; tornadoes; blizzards; floods; and others'. Many of these regulations have recently been reviewed. For example, the Senate is urging the Secretary of Homeland Security to promote the National Fire Protection Association standard on Disaster/Emergency Management and Business Continuity. This requires that the owners and managers of public buildings conduct 'hazard identification and risk assessment'. The aim is to provide the best means of "instructing occupants to evacuate the building or shelter in place"(NFPA, 2005).

European legislation is also intended to ensure the prompt evacuation of public buildings. Directives, 89/391/EEC and 89/654/EEC, describe minimum standards that should be enforced by legislation in each member state. The UK Fire Precautions (Workplace) Regulations were amended in 1999 to meet these directives. All occupants must be alerted and leave buildings safely in the event of a fire. Employers are responsible for the outcome of any adverse event. The focus of the UK amendment was also to introduce a *risk-based approach* to fire regulations. Building owners and managers must demonstrate that any precautions are appropriate to the likelihood and consequences of any hazard. Evacuation measures could be use to demonstrate mitigation of the potential consequences of an adverse event.

This risk-based approach has been adopted within the provisions that guide the use and management of large public buildings within particular domains. For instance, the 2001 Department of Health guidance covering Scots hospitals includes requirements that "NHS Trusts must have an effective fire safety management system which provides for…means of ensuring emergency evacuation procedures for all areas…means of ensuring that procedures are in place to undertake fire risk assessments throughout the Trust and to monitor these on a regular basis". Individual NHS Trusts must also appoint specialist Fire Officers who can provide technical support and "involvement with estates staff and others, in fire safety audit and fire risk assessments and assisting with reports to management" (NHS, 2001).

Most recent legislation advocates the use of risk assessment to identify the hazards that threaten the safety of public buildings. The development of evacuation plans and the provision of escape routes provide owners and managers with means of mitigating the risk of fire etc. The following sections argue that a risk-based approach should be extended beyond the immediate *causes* of an evacuation to consider the particular hazards that might *prevent* occupants from escaping a building. The evacuation of the World Trade Center has shown us that the owners and managers of large public buildings must consider the possibility that some emergency exits are blocked and damaged whilst others remain open (Johnson, 2005). They must also consider what might happen if it is no longer possible to use the public address systems that are often used to initiate evacuations.

1.3 Overview of the Paper and the Proposed Approach

There is very little practical advice on how to adopt the risk-based approach that has been advocated in Europe and the USA. The owners and operators of large public buildings continue to relay on subjective inspections and walkthroughs both to assess the risks that can lead to an evacuation, such as a fire hazard, and also the integrity of evacuation routes. These informal techniques have been widely criticized in the aftermath of major fires (Johnson, 2005). It can also be difficult to adapt more objective forms of risk assessment to represent and reason about the risks that might complicate the evacuation of large public buildings. Section 2 will show how the gates within a fault tree can be used to identify the conjunctions and disjunctions of basic events to represent the ways in which bottlenecks can arise through poor building design and fire damage or barriers created by temporary structures and partition walls. However, these techniques must be supported by evidence from previous fires and live drills if they are to account for the wide range of human behaviors that have been seen in many evacuations. The lack of national and international databases for evacuation information, especially about the mass of near miss and low severity incidents, restricts the insights that can be obtained from previous incidents. Ethical and practical considerations limit the use of 'live' evacuation drills and exercises. For instance, it can be difficult to conduct these drills in institutions such as banks and hospitals that are intended to provide 24/7 services.

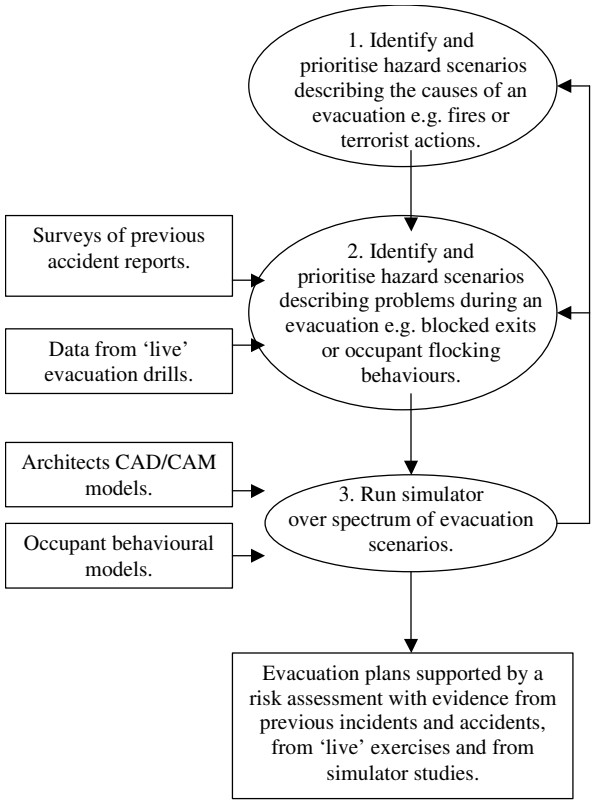


Figure 1: Overview of the Approach and Structure of the Paper

This paper argues that computer-based evacuation simulations can be used to supplement live exercises and more conventional risk-assessment techniques. Figure 1 provides an overview of the proposed approach and also sketches the structure of the argument in this paper. As can be seen, the approach begins with a risk assessment, as recommended by the legislation reviewed in the previous section, into fire and other hazards, such as chemical release, that might cause an evacuation. Part of this process will include some consideration of the ways in which improved evacuation procedures will help to mitigate the risks. The output from such an initial analysis can then be used to inform a risk-based approach to evacuation management.

The second stage of the risk based approach to evacuation management, illustrated in Figure 1, uses existing risk assessment techniques, including Fault Trees and FMECA, to map out the ways in which an evacuation may fail. The intention is to identify the most critical hazards, in terms of consequence and likelihood that could prevent egress from large public buildings. It is important to reiterate the difference between this stage and the previous phase that involves more a conventional assessment of the events that trigger an evacuation. For example, occupants can be forced to leave a building from a fire or from terrorist action. However, their evacuation might be impeded in both cases by the inadequate lighting of internal stairwells or by occupant flocking behaviors. This second stage, evacuation risk assessment must be informed by an analysis of previous situations where occupants have been forced to leave similar buildings using accident and incident reports. The objective of this analysis is to identify critical hazard scenarios that will then be the focus for further investigation using software simulation. The analysis can also be informed by insights from 'live' drills, although this may not be possible in new buildings.

The third stage is to develop and run interactive simulations for the building and occupant population being considered. Subsequent sections of this paper will describe a suite of tools that automatically derive these simulations from the CAD/CAM files used by architects. This reduces the costs associated with simulation and also opens the potential to run evacuation simulations before a building is constructed. This simulation stage also relies upon behavioral models for the building occupants. Young and assertive

individuals will often respond quite differently to, for example, large family groups during emergency evacuations.

Figure 1 describes an iterative approach to evacuation management. Simulations can be shown to many different stakeholders, including building occupants and emergency personnel. These consultations often yield large numbers of additional hazards and evacuation scenarios that must be integrated with any existing risk assessments conducted in the first and second stages. Similarly, annual or monthly evacuation exercises can yield further insights that must be incorporated into the evacuation planning process.

The evacuation of a large, general hospital will be used to illustrate the application of the techniques summarized in Figure 1. The number of fires that occur in hospitals each year justifies this decision. For instance, there are approximately 2,500 major fires in Scots hospitals alone. In the United States, there are 3,500-4,000 fires involving multiple fatalities in nursing and assisted living homes per annum. No accurate records are kept for the number of incidents that lead to the deaths of single individuals. The focus on hospital evacuations is also justified by public concern following particular incidents. The Seacliff Mental Hospital Fire in New Zealand continues to have an impact on the planning of healthcare institutions in that country and remains one of the worst single incidents in their history with thirty-seven deaths. In 2003, 30 patients died in a hospital fire in Belarus while another 10 died in a fire at the Greenwood Health Care Center in Cennecticut, USA. The January 2004 Rosepark Care Home fire in Uddingston killed ten patients and sparked a national debate on the safety of healthcare institutions in the UK. As I write this paper, news has arrived of 17 deaths in a hospital fire in Costa Rica. Public concern is justified even when there are no direct fatalities. For example, a recent arson attack on London's University College Hospital cut off oxygen and power supplies and forced a partial evacuation that placed patients and staff at risk. These events motivated a roundtable into Healthcare Fire Safety, held by the International Association of Fire Chiefs (IAFC, 2004).

Risk-based approaches to evacuation planning pose significant challenges for large hospitals. Many of these institutions rely on a mixture of legacy buildings together with more modern facilities. Further complexity stems from the diversity of patients who are treated in many healthcare facilities. These can include ambulatory outpatients as well as individuals who rely on wheel chairs. It also includes patients who cannot be moved from their beds or who can be moved but only after their care has been transferred to a complex array of mobile monitoring and treatment devices. Complexity also arises from the range of detailed procedures that hospital staff use to ensure that patients are evacuated away from a hazard as soon as possible.

2. Identifying and Prioritizing Evacuation Scenarios

At present most managers and owners identify the hazards that might lead to an evacuation or prevent it from being completed by informal walkthroughs with designated Fire Safety Officers. Paper-based forms provide check boxes to note the presence of particular hazards within a building. For example, these are often used to indicate the obstruction of fire escape routes by non-permanent objects or to indicate the need for additional fire extinguishers. Informal 'walk throughs' are far from ideal. Confirmation bias occurs when inspectors consistently identify the presence of particular hazards but also consistently miss other hazards when they work together. Organizational bias occurs when the managers and operators of a building act to influence the outcome of a walkthrough by promising actions, such as the removal of obstacles, before a report is published. Individual bias occurs when inspectors promote particular concerns beyond the level that might otherwise be justified for a particular hazard. Many of these problems remain hidden until an evacuation occurs because the judgments made during a fire inspection are not usually supported by detailed evidence from previous fires or evacuation exercises. The gradual introduction of the risk-based approach has also created a situation in the UK and in the US where managers have introduced rolling-plans of inspection across large portfolios of buildings. Changes in building occupancy create a continual need to go back and re-inspect areas that were considered only a short time before. This can lead to further disagreement where practices that were safe in a previous inspection may no longer be acceptable under new operating conditions.

A number of groups have advocated risk assessment to counter the perceived weaknesses of unstructured, techniques for analysing fire hazards in public buildings (Chamberlain, Modarres, Mowrer 2002, US National Fire Protection Association, 2004). However, most previous research focuses on subjective risk assessment for the events that trigger evacuations, such as fires or terrorist action. There is little quantitative work on assessing the risk of different evacuation scenarios. Existing techniques, such as FMECA or Fault Trees, could be used. Table 1 illustrates the FMECA approach using column headings based on those in US Military Standard Mil-Std-1629A. As can be seen, analysts must identify the various sub-systems that support an evacuation. They must then identify the causes of the various failures that can affect these systems. For example, a sprinkler system can help evacuations by reducing smoke levels and can buy additional time for an evacuation by limiting the growth of a fire. Such support can be jeopardized if the aprinklers' water supply blocked.

Evacuation of Area 1: Treatment Rooms							
Ref	System/ Equipment Failure	Cause	Effect	Detection	Mitigation/ Compensation/ Safeguards	Overall assessment	Overall criticality
1A	Sprinkler system	1. Blocked	Water cannot be discharged through system.	Pressure diagnostic tests.	Clean system using steam/pressure. Possibly consider redundancy.	Sprinkler system failure from evacuation perspective may prevent clearing of smoke and decrease time available for evacuation.	B.
2A	Evacuation corridor	1. Bottleneck caused by trip or fall.	Stampede and possible crush injuries.	Fire officers monitor egress of personnel and patients from all areas.	Review evacuation routes. Ensure supervision of egress at key points offering assistance to some occupants.	Critical in areas where many occupants meet at same time, eg stairwells & landings.	A.

Table 1: Example FMECA for a Hospital Evacuation

Table 1 also illustrates some of the changes that must be made if FMECA is to be used to consider the wider hazards that can arise during an evacuation. The final row considers the problem of a bottleneck in an evacuation corridor that can be caused when occupants stumble and fall during an emergency. It would be unusual to consider a corridor as a 'system' within other forms of FMECA. However, the application of the approach to building evacuations forces the analyst to consider the layout and operation of such escape routes as a primary concern. A number of issues remain. For instance, Table 1 also includes a criticality assessment. The product of likelihood and consequence determines this in the usual manner. However, any assessment of these two factors depends upon a large range of different environmental and contextual factors. The likelihood would depend upon the number of people in the building. It would also depend upon their distribution and their average speed. These issues, in turn, determine whether large numbers of people will reach any particular bottleneck at the same time. The severity of any consequences depend on

a similar broad range of factors such as the age and physiological condition of the occupants, the speed they were travelling, whether they were panicking, whether there was smoke etc. One approach would be to associate the most plausible worst-case criticality with each row in an FMECA evacuation table. In order for this approach to contribute to future evacuations it is important to identify those locations in a building where the 'plausible worst-case scenarios' are likely to occur. This would then enable managers and occupiers to re-design the layout of evacuation routes or, for instance, to deploy additional fire officers.

Fault Trees can be used to focus more on the likelihood of evacuation hazards. Each of the causal factors in Table 1 could be considered within the disjunctions and conjunctions of such diagrams. This approach offers a number of advantages because inspectors can use the resulting diagrams to explain the reasons why they are concerned about particular hazards during an evacuation. Figure 1 illustrates an evacuation Fault Tree. As can be seen, crush injuries can occur given that a building occupant falls to the floor and they are in an 'at risk' group, such as the elderly. Such falls can occur if an evacuation route is obstructed or the visibility is poor. Fault tree diagrams can also be used to identify appropriate mitigation techniques for each of the factors that contribute to the likelihood of an evacuation hazard. Building managers might provide additional emergency lighting, luminous handrails and step indicators in areas where smoke accumulates. Additional fire officers might also be recruited to guide 'at risk' residents to appropriate exits.

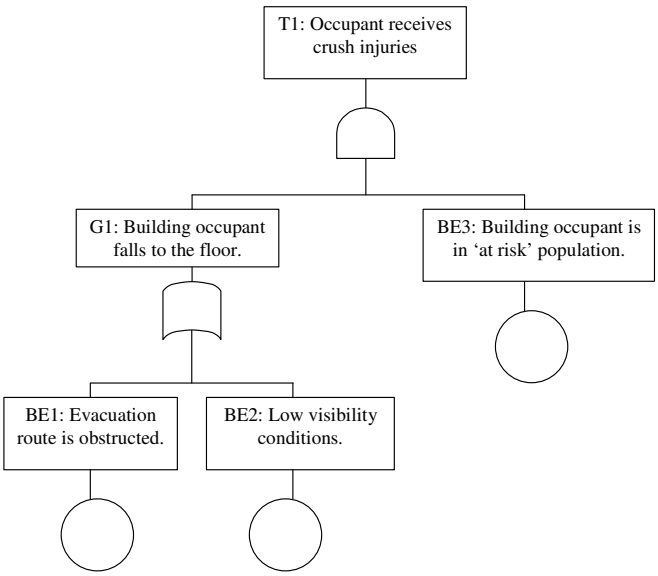


Figure 1: Example of an Evacuation Fault Tree

The application of Fault Trees to support a risk-based approach to the evacuation of public buildings raises a host of further questions. For example, many of the most powerful applications of Fault Tree analysis rely on the propagation of failure probabilities through the tree to help calculate the likelihood of a top-level adverse event. This can be fairly straightforward for some events. For instance, building investigators can survey the population of occupants to determine the likelihood of an individual being 'at risk' of severe injuries during a fall. Previous studies of different fires can also be used together with an analysis of building contents to estimate the likelihood that a fire might result in low visibility conditions. However, it is unclear how to calculate the likelihood of an 'evacuation route being obstructed' within a hospital given that such obstructions continue to occur even though regular inspections are conducted and procedures are drafted to avoid such hazards. Many of the organisational and individual biases that affect ad hoc walkthroughs will also influence attempts to obtain evidence for the likelihood estimates in evacuation Fault Trees.

The key issue here is that most existing risk assessment techniques provide a high-level structure or template for arguments about the risks associated with particular hazards. They do not provide a panacea for the host of more detailed problems that arise when conducting a risk-based approach to evacuation.

These techniques are useful because they provide analysts with a high-level means of identifying important hazards, including obstructions and reduced visibility. They cannot easily be used to assess the likelihood and consequences of relatively small changes to the geometry or functions conducted in areas within complex public spaces.

2.1 Insights from Previous Accident and Incident Reports

It is important that the managers and operators of large public buildings learn as much as possible from previous adverse events. For example, it is relatively uncommon to witness panic. Disbelief is a more frequent response to an initial warning about an adverse event. Occupants often attempt to establish the credibility of a warning by asking colleagues or members of staff (Bryan, 1982). There is also a tendency to ignore any warning if there is conflict or ambiguity. For example, building occupants will delay an evacuation if an audible alarm is not located within their immediate vicinity. Such general findings can be confirmed by specific investigations into previous hospital fires. Edelman et al (1980) analyzed the evacuation of a nursing home and stressed the impact that previous false alarms had upon occupant behavior. The alarms were ignored until several patients began to shout 'fire'. Only one psychiatric patient showed symptoms of panic during the evacuation. There are further common factors between hospital evacuations and emergency response in other buildings. Proulx (2001) describes how many people ignored fire exit signs and rushed back in the direction that they had used entered a terminal at Munich Airport. Similarly, two people were killed in the evacuation of the Lowenbrauskeller when the majority of occupants walked part 8 emergency exits to reach the main entrance. The 2003 fire in Rhode Island's Station nightclub provides a further example. Most of the 300 customers retraced their steps back to the main exit. Those who reached this area had to force their way through a bottleneck created by a ticket booth leading to numerous crush injuries. Edleman et al (1980) describe a similar evacuation strategy for the staff in the care home fire. 95% (85) of the patients on the affected floor were led down a single staircase even though three others were available. This staircase was the normal route used by staff and patients between the two floors. The other three were evacuation routes and were fitted with entry alarms, hence there was a reluctance to use them even when the fire justified this. In consequence, the evacuation took longer than expected by the building designer and by the Fire Officers who were involved in the certification of the building.

Reports into the causes of hospital evacuation are published in two formats; aggregated information about minor incidents and detailed reports into major investigations of single adverse events. A recent US Food and Drugs Administration report into fires involving electrically powered hospital beds can illustrate aggregated information. The likelihood of any individual hospital experiencing one of these fires is relatively low. However, regulatory agencies such as the FDA collect this information in order to ensure that lessons learned in one organization can be passed to others. Their records revealed that this hazard accounted for over 100 fires in less than ten years. Approximately 25% of the reports failed to identify any particular cause for the smoke or flames that were observed. The remaining 75% were due to motors overheating, overheating of bed capacitors, arcing at the plug and wall plate due to poor fit, plug damage etc.

Aggregate studies of previous incidents are useful because they provide insights into trends that can only be detected as a regional or national level. However, they are typically targeted at the causes of fires and rarely yield specific insights into particular evacuation procedures. This information is, typically, easier to extract for more detailed reports into individual adverse events. For example, the US National Fire Protection Association (1993) has published a summary report into a hospital fire in Booklyn. This illustrates the evacuation problems that arise when fires are triggered by causes similar to those described in the FDA aggregate report, cited above. In this case, a fire quickly ruptured the oxygen hoses that were being used to treat a patient. The hoses were directly attached to wall outlets and the resulting free-flow of oxygen fed the resulting fire. Large amounts of smoke were forced into the hall and throughout the patient floor. A relatively small fire, therefore, escalated far more quickly than might otherwise have been the case. It also forced the evacuation of larger numbers of people than might have been expected in a residential setting given the close proximity of large numbers of bed-bound patients within the hospital wards.

This example illustrates the complex nature of hospital evacuations that force managers and staff to make detailed plans for the various scenarios that have been mentioned in previous sections. Nurses and Fire Officers may have to delay the evacuation of patients in order to find the time necessary to prevent a fire from spreading. In this case, staff may be diverted from evacuations to close the pipeline zone valves that control oxygen enriched treatments. The level of detail that is necessary in evacuation scenarios can also be illustrated by this example. Nursing staff must consider the consequences if they close the valves that control the oxygen flow to patient's rooms. Such actions will reduce the amount of oxygen feeding a fire; it will also cut off the oxygen supply to other patients within the affected zone. Residual pressure in the pipeline will often allow a short interval before a patient's treatment will stop completely. This provides nursing staff with the opportunity to make alternate arrangements, for instance using bottled oxygen supplies. However, these also create hazards if they are stored on floors where a fire has been detected.

A similar fire caused by smoking materials in a patient's bed led to the deaths of five patients in a Virginia hospital (NFPA, 1994). This incident is typical of many in well-prepared hospitals. The building itself was constructed from fire-resistive materials. Hospital staff had also been well trained to respond to such emergencies. However, smoke spread into concealed spaces about the ceilings of the patients' rooms and several factors combined to prevent a prompt evacuation. These can be summarised as follows:

- Delayed fire discovery. Staffing levels drop at night and this can increase delays in detection. Staff also will often be preoccupied with other tasks. Many hospitals, especially in legacy buildings have areas in which fires can break out, such as linen closets and equipment stores. Many of these areas cannot easily be covered by accurate fire detection technology and are not easily inspected by busy clinical staff.

- Delayed communication with emergency services. The system connecting the hospital alarm to the local fire department had been taken out of service. Such equipment problems are typical in many health related organisations where direct patient care is often seen as the primary objective and issues such as fire safety are paradoxically seen as having a secondary importance.

- Oxygen enriched environment. The severity of the fire when it was discovered and the rapid development of untenable conditions. The Virginia hospital fire is typical in that many hospital fires rapidly develop to threaten the safety of large numbers of patients. The role of oxygen and other volatile gases has been mentioned above. In addition, many of the doors that connect patient rooms and wards to corridors are deliberately left open. This can occur even for fire doors. Wedges can be used to help patients call for attention from busy nursing staff. Open doors assist ventilation in legacy buildings. Door can also be wedged open by busy staff as they clean rooms or distribute equipment and supplies.

- Complex building design. Hidden areas between individual rooms helped to propagate fire and smoke. Again, this is typical of many legacy buildings where, for example, false ceilings have been introduced into Victorian hospitals. Ventilation and cabling ducts can also introduce hidden transmission routes. It is important not to underestimate the impact of such passageways. In this Virginia fire, one patient died far away from the seat of the fire while many others survived. Such 'hidden' transmission routes may also force staff to consider evacuating areas that are well beyond the immediate vicinity of a fire.

- Lack of sprinkler system. Finally, the report into this incident criticised the lack of a sprinkler system in the room where the fire began. Such systems delay propagation and buy extra time during an evacuation. However, as with many other aspects of hospital evacuation, there are cost-benefit trade-offs if a sprinkler system is used when patients rely on sensitive electrical equipment to provide vital support.

The Virginia incident illustrates how reports into previous hospital fires can be used to guide the identification of evacuation scenario that other hospitals use during drills and exercises. These reports can also be used to identify the likely consequences and hence provide direct evidence in support of particular risk assessments. However, hospitals are extremely complex buildings. The hazards vary according to the layout and function of different areas. It is, therefore, critical that Fire Officers consider a broad range of

reports rather than attempting to generalise too widely from a narrow range of examples such as the Virginia and Uddingston incidents. For example, it is far easier to initiate the evacuation of patients from their rooms and wards than it is to respond to fires in an operating theatre. The US Joint Commission on Accreditation of Healthcare Organizations (2003) estimates that there are 100-200 of these fires in the US each year. The risks of fire again include an oxygen-enriched environment with a wide range of possible ignition sources including lasers and cautery units. However, evacuation can be hazardous both for staff and for patients who typically require intensive care whilst under sedation. Specialist training is required in order to use hand-held fire extinguishers and fire blankets in sterile environments.

The way in which the JCAHO have to estimate the number of surgical fires in the US raises another important issue; there are no national or Federal registers that provide a central record for most of these events. In Scotland, for example, it is a requirement that all National Health Service organisations report fires involving death or serious injury to the Health and Safety Executive. They must also report fires involving death, serious injury or damage on a large scale, to the Department of Health. This focus on relatively serious incidents limits feedback on less serious events that can provide insights into successful evacuation techniques. Between 1994-2001 only 6 reports were made. 5 involved patients smoking and 1 involved 'willful' fire raising (NHS, 2001). It can also be difficult to access information about more serious events; which are often subject to litigation. In consequence, Fire Officers rely on 'war stories' and informal anecdotes that are passed by word of mouth during periodic meetings and evacuation training sessions. This contrasts strongly with the legal reporting requirements that govern the failure of the devices that cause fires in healthcare settings.

2.2 Insights from Evacuation Drills

Many evacuations in response to minor incidents and false alarms are never reported. In the absence of suitable national and international exchange mechanisms, analysts must rely on live drills and exercises to provide insights into their evacuation strategies. These exercises also play an important training role by providing staff with an opportunity to rehearse and coordinate their response to an adverse event. This creates a circular problem. Drills are used to identify potential problems in an evacuation. However, it is important to anticipate potential problems that can arise during an evacuation so that they are scripted in such a way that staff are challenged to respond to these problems. In consequence, many organizations with a strong safety culture will use evacuation drills in an iterative manner. Subsequent exercises are designed to test weaknesses that have been exposed in previous drills (Johnson, 2005).

It is important to illustrate the scale and complexity of evacuation exercises in hospitals. For example, a US hospital recently conducted 3 mock fire drills during a 6-week period. One scenario started when the tip of an electrosurgical pencil that had not been placed in a holster ignited a drape or cover (McCarthy and Gaucher, 2004). Staff members rapidly removed the cover from the patient by throwing it on the floor and using a fire extinguisher. Other colleagues were informed of the fire. At this point, however, the staff running the simulation intervened to inform them that the fire had spread. A senior nurse began to coordinate the evacuation of operating room staff. There was initial confusion about the best way to transport the patients to a triage point. Partly as a result of this several adjacent rooms were evacuated at the same time causing temporary gridlock in the corridors. This evacuation drill simulated the movement of intubated patients using the operating room bed with a bag-valve mask. The exercise also required staff to move individuals with open incisions. Wounds were packed with sterile, saline-soaked laparotomy sponges and then covered with sterile drapes. The evacuation scenarios were also scripted to determine whether staff knew which items of equipment needed to be evacuated with their patients. They had to collect enough instruments to close the incision even though the evacuation plans provided for sterile equipment to be available in the triage area. Staff were also supposed to know that it was not necessary to transport the anaesthesia machine with the patient.

Debriefing sessions were held after each exercise and enabled staff to provide additional information about a wide range of problems. Evacuations did not always proceed in an orderly fashion. Some staff were unsure about how to use a check sheet describing the key tasks for coordinating an emergency response. There were delays in calling for backup when both the patient and the anaesthetist were 'injured' during the exercise. Debrief sessions also helped to identify problems that were not always visible to the organisers. For instance, one anaesthetist said that they would have evacuated a patient using the back door of the

operating theatre. This exit opened onto a steep incline above a busy road. The hospital was then able to respond by posting additional guidance to staff in that area, including signs on the doors that discouraged their use as an evacuation route.

These exercises also provided information on more 'systemic' problems. For example, the hospital paging system played a central role in coordinating the emergency response. During the exercises, it emerged that many announcements could not be heard. Staff then had to either contact the desk issuing the calls or leave their posts to seek further clarification. It also emerged that no one was sure what would happen if it were to be damaged. As a result of these exercises, changes were made in the way that messages were sent around the hospital. A messenger position was opened and plans were made to distribute walkie-talkies in case the existing communications infrastructure was compromised during an adverse event.

Evacuation Procedures in the Case Study Hospital

These exercises provide staff with the opportunity to practice complex evacuation procedures. For example, the hospital that forms the case study in this paper exploits 'horizontal evacuation'. Staff move patients from a hazardous area to a place of safety on the same floor, for instance behind fire resistant doors and walls. Only if the situation worsens significantly will they consider moving patients to other floors and eventually out of the building entirely. The evacuation follows a predetermined plan in which staff must first locate the source of any hazard and then ensure that the proposed destination will keep them free from any immediate danger until the emergency services can arrive. This implies that the destination must be more secure that the area from which a patient is being moved. It is also important to continue to ensure that there is a protected route from the place of safety to an exit from the building. Different classes of occupant raise different concerns during an evacuation. Patients in immediate danger must be moved first. Some assessment may have to be made about whether the risk of moving the patient is greater than the risk posed by the fire or other hazard. Non-ambulatory patients can, typically, be considered before ambulatory patients and visitors. Wheel chair patients are grouped together and then taken to a place of safety by teams of nursing staff. Staff can lead groups of more mobile patients to safety in a single journey. Patients must be taken to a place of safety that does not impede the ingress of emergency personnel. This is important because there is a danger of injury as equipment and people move in to tackle a fire or similar hazard.

Even this superficial description should illustrate the additional complexity that such evacuations can pose beyond the normal workplace drills that most people will be familiar with. However, these drills can be vital in gathering information about the time that is required in order to complete an evacuation. For example, each ward in the hospital appoints one person to coordinate the evacuation. Their performance can vary widely according to the level of staffing and the mix of patients they have to care for. Drills have shown that it takes three people around five minutes to disconnect patients from fixed equipment and reconnect them to mobile monitoring units etc. It can take up to fifteen minutes to transfer a conscious patient from a bed into a wheelchair. Once patients are ready to be moved, drills provide further information about the time required to evacuate them to a place of safety. For example, in most floors in our case study hospital it is possible to find refuge within approximately twenty meters of each patient's room. On average it takes staff seventy seconds to move a patient from various locations within their room to a place of safety. It takes a further thirty seconds for staff to return to the patient's room to collect someone else. This would occur if several wheelchair patients have been grouped together for evacuation.

Previous sections have argued that there is a great need for healthcare institutions to share insights provided from previous evacuations. However, the utility of this information is limited because evacuation procedures vary between healthcare institutions. In particular, different patient profiles will influence the evacuation techniques that are used. For example, Wisconsin like many other US states urges staff not to use the 'horizontal' evacuation techniques described for the case study hospital when evacuating 'Intermediate Care Facilities serving persons with Mental Retardation'. Evacuations should move all patients outside the building; 'this is required, regardless of building construction certification' and such a facility 'may not use defend in place methodologies' even during evacuation drills.

Limitations of Live Drills and Exercises

As mentioned, 'live' evacuation drills serve a double purpose. They can be used to establish that minimum evacuation times continue to be met. This is important because fire exits can be inadvertently locked or obstructed. Fire drills can also be used to ensure that occupants are familiar with necessary evacuation procedures and routes. Hence, in many countries it is a requirement that these drills be performed on a regular basis even after it has been demonstrated that a building meets the initial regulatory requirements, described in the previous sections. For example, many US hospitals conduct exercises in key departments at least once every three months in order to meet insurance requirements. This creates a host of practical problems. For instance, most exercises are conducted during the day. However, it is equally important to provide night staff with an opportunity to practice their skills and also to observe the impact that evacuation procedures have at such times. The results of these night drills can often be very different compared to the same patient population during the day. Many more patients require assistance after being roused from sleep, especially if they are under sedation. Staffing levels are often reduced at night and so coordination can become far more problematic. Many hospitals rely on a greater proportion of agency and part-time staff at night. It can be difficult to ensure that these temporary staff members are familiar with evacuation procedures. Some of these issues persuaded the Department of Health in Scotland to change its regulations and "reduce the need for annual fire safety training for all staff where a full risk assessment has been carried out" (NHS, 2001). However, NHS Trusts must ensure that "procedures are in place within the Trust to provide regular fire safety training for all staff, appropriate to the duties of the staff and their place of work" and provide "means of ensuring that appropriate training exercises are undertaken at least annually for the fire response teams and other staff who are involved in patient evacuation". There are, however, a number of limitations that affect the utility of 'live' fire drills as a means of assessing occupant's ability to escape from large public buildings, such as hospitals:

1. **Sustained Costs.** For many employees, fire drills are little more than a nuisance every month. However, there are considerable costs associated with evacuation drills in hospitals. They can have knock-on effects that disrupt complex healthcare schedules, including surgical lists. It is for this reason that the Scots regulations, cited above, advocate that a risk assessment be used to determine those personnel who must be involved in an annual evacuation drill.

2. **Limited Accuracy.** It can be hard to use fire drills to simulate a range of potential hazards. There is a tendency to simply ensure that everyone in the building knows where the nearest exits are located. Few drills determine the impact of forcing occupants to find alternate forms of egress should these become blocked during an incident. Previous studies of evacuations within other healthcare institutions, including long term residential care, have shown that periodic drills only have a limited effect in persuading staff and patients to use fire exits rather than the main entrances for a building. Similarly, many exercises do not involve the participation of external agencies who may be required to enter the building to complete an evacuation.

3. **Short 'Shelf Life'.** Changes in building use affect the results from 'live' simulations, especially for hospitals that rely on annual drills. In the meantime, large items of furniture such as filing cabinets and beds, as well as other items of clinical equipment can accumulate in areas that obstruct horizontal evacuation procedures. Given the day to day demands on many healthcare institutions it can be difficult for staff to remember that they may have to move several beds and wheel chairs down smoke filled corridors within a short interval after an evacuation has been ordered. In consequence, a successful drill in the immediate past can provide only limited assurance of a successful evacuation in the future. The limited 'shelf life' of evacuation drills is also affected by the rotation systems that govern the operation of many healthcare organizations. For example, anesthetists may work in many different departments across several different hospitals. Operating theatre staff work in rotation. Hence, fire drills that involve specific teams may have to be repeated to involve a broad cross-section of the individuals who may be called upon to act together in an emergency.

4. **Lack of Design Focus**. It is difficult to use the insights from evacuation drills to inform the design of large public buildings. For example, the UK NHS has been involved in the construction of several large, centralized hospitals such as the New Gloucestershire Royal Hospital. This must satisfy design criteria that bring conflicts of interest in terms of acoustic performance, ventilation

and comfort whilst also meeting evacuation provisions in the national Fire Codes. Drills cannot easily be conducted to provide insights into evacuation times for buildings that do not yet exist. Instead, architects and managers must focus on a narrow set of 'static' factors such as the size and location of emergency exits. They cannot easily account for the distribution of semi-permanent obstacles or even the detailed staffing levels throughout the working day that have a profound impact on an evacuation. It would be useful to have a system that designers might use on an iterative basis to assess the effects that changes might have as they revise the layout and structure of a potential building.

5. **Danger.** Several firefighters die in evacuation exercises each year, either form 'workplace accidents' or from existing medical conditions. In consequence, extreme care must be combined with appropriate risk assessments before such trials can be attempted. There are additional ethical and legal complications when subjects may be drawn from the potential occupants of a building. In consequence, restrictions can be placed upon a healthcare organization's ability to involve patients in these exercises. Informed consent is a prerequisite. It can be difficult to obtain sufficient support from patients whose primary concerns do not focus on their involvement in a drill. Many US states follow the Pennsylvania code in letting healthcare institutions decide whether or not to involve patients;

6. **Poor Reliability.** If the same exercise is performed on several different occasions within a limited period of time then the outcomes can be very different. Contextual factors have a profound impact upon evacuation rates. For instance, if an individual begins a prompt evacuation then their peers will often follow shortly behind. However, if individuals delay their initial evacuation to complete particular tasks, such as closing down a computer workstation, then others in the group will often feel the need to do the same before beginning to egress from the building. Such dynamics of group interaction reduce the reliability of results obtained from specific evacuation drills. In hospitals, evacuations must often be coordinated by a small number of key individuals. If those individuals forget to alert all of their colleagues or skip necessary steps in an evacuation plan then the outcomes can be significantly affected, as illustrated by the drills mentioned in previous paragraphs.

Computer-based simulation tools address some of the limitations of 'live' exercises. For example, it is possible to explore what might happen by altering the layout of a building before it is constructed. Managers can simulate the effects of different staffing levels on average evacuation times. Similarly, they can explore the effects of increasing patient numbers or altering the mix of patient conditions being treated within a particular area of the hospital. It is possible to interactive block escape routes as the simulation progresses. This software has a variety of potential end users from architects through to Fire Safety Officers. Regulatory agencies, certification bodies or the emergency services can use them during the approval process that is required before a building can be opened for operation or approved for construction. Occupiers can also use these tools to examine the potential impact of changes in the architecture or operation of a structure. The results from previous exercises can be used to calibrate the findings from these models, which also avoid many of the costs and risks associated with exercises involving real patients. The following section, therefore, introduces some of the design challenges that arise during the development of such software simulations.

3. Simulating a Spectrum of Evacuation Scenarios

There have been a number of previous attempts to develop computer-based simulations of evacuation behavior from large public buildings. For instance, the UK Atomic Energy Authority (2002) has developed the Egress simulator. This tool enables users to draw a simple floor plan of the building under investigation. Hexagonal cells are then used to segment the area. Different types of cell are used to distinguish between internal walls, between areas that are already occupied by people and movable obstacles such as tables and chairs. The Fire Research Service adopts a more elaborate approach (BRE, 2004). CRISP users can associate behaviours with each occupant. These are described in terms of actions, which may be abandoned, and substituted by new ones in response to changes in their environment. Individuals can also investigate, warn others etc. before starting their evacuation. In contrast to Egress and CRISP, the EXODUS system has also been adapted for use in the aviation and maritime

industries (Owen, Galea and Lawrence, 1996). Key attributes of the behavioural modelling include the ability to dynamically insert individuals during a simulation. The EXODUS tool provides important facilities in terms of signage annotations so the end-users can simulate the impact of providing additional warning and information notices. Dynamic behaviours can be altered so that individuals will automatically seek alternatives if they see that a particular exit is already congested.

Evacuation software, typically, relies upon models of human behaviour to drive their simulations. For example, we have already mentioned previous incidents in which occupants have first tried to establish the credibility of an alarm before starting to move away from a potential hazard and towards a place of safety (Bryan, 1982). Simulations can mimic these findings by introducing a fixed delay into each run. However, more elaborate models can also be developed to consider a range of more detailed factors that can influence this delay before evacuation. These include the perceived threat posed by the alarm, the degree of preoccupation with the task to hand, familiarity with evacuation procedures from previous drills etc. It is also important to consider the social and team factors that have been shown to influence evacuation times in 'live' hospital exercises. The Federal Emergency Management Agencies have argued that the stronger the bond between group members, the more likely it is that one member will put their own life at risk to protect another group member. Tong and Carter (1985) describe a further form of social behaviour that occurs as crowds grow and groups converge. "Flocking" can attract more people into areas that are already crowded. This form of behaviour can act as a catalyst to flight. Personality traits such as assertiveness have been shown to influence decision-making and behaviour under stress. For example, the Transport Canada Personality Profile 2 (TCPP2) identifies 13 characteristics that influence behaviour during evacuations. Projections based on the results of their experimental studies suggest that 20% of people are 'highly assertive' or 'goal directed'. These individuals can have evacuation times that are up to 25% faster than the 15-18% of people who are classified as being in less goal-oriented groups (Latman, 2004).

Not only must evacuation simulators consider social and cognitive characteristics, they must also account for different physiologies. Age and physical limitations determine the speeds at which people will travel through the building during an evacuation. However, these characteristics cannot be viewed in isolation; a panicking individual is more likely to travel at greater speed than a person who is calm. In the GES tool, each person is assigned an initial speed. The medium speed is set to be 1.4 ms^{-1} (Thompson and Marchant, 1995). The low and high-speed groups are set to have a pace that is 80 and 120 percent of this respectively. These values can be set by the user to calibrate their system. However, these initial values are based on empirical observations that take into account individual pace under different crowd densities. This preferred walking speed of evacuation is sustained unless they cannot make any further progress because one or more people in front of them blocks their path.

We have developed the Glasgow Evacuation Simulator (GES). This tool relies on Monte Carlo techniques to introduce non-deterministic behaviour into scenarios. Random numbers are generated and then compared against probability distributions to help simulate individual and group behaviours. This ensures that building occupants do not always follow the same course of action during each run of the simulation. They are, however, more likely to perform those actions that are considered to be most probable during an evacuation. The probability of particular behaviours can be directly informed by previous incident reports and by the observations derived from evacuation exercises. In consequence, it supports the iterative approach to fire and evacuation risk assessment illustrated in Figure 1. It is informed by rather than being a substitute for 'live' drills. One innovative feature of the GES is that it uses the 3D models that can be obtained from architects' design tools. Unlike many other simulators, there is no need to build specialized models for the evacuation simulator. This reduces costs and allows a tight integration between the simulator and the design of such structures. As shown in Figure 1, the ability to derive simulations from the files of tools such as AutoCAD enables us to simulate buildings that have yet to be constructed.

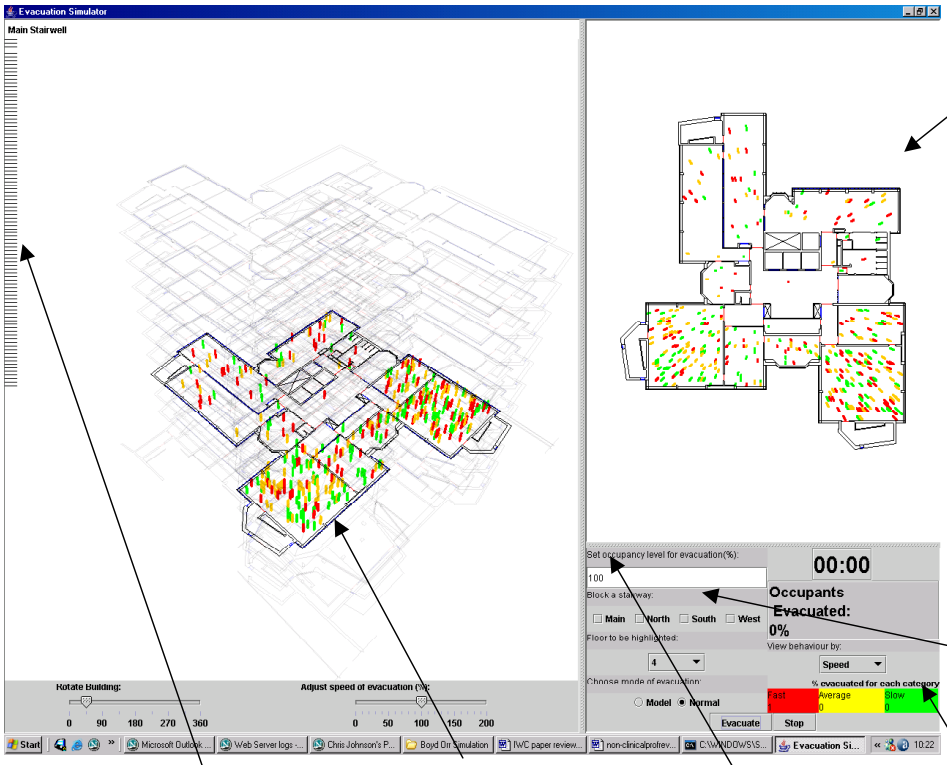


Figure 3: User Interface to the Glasgow Evacuation Simulator (GES)

Figure 3 illustrates the application of the GES to model evacuations from a large auditorium complex within the Boyd Orr building in Glasgow. As can be seen, the interface enables users to vary the occupancy levels in the building. Users can also interactively open and close emergency exits as a simulation progresses to model the effects of damage to the building or intervention from the emergency services. It is also possible to specify whether users will follow a 'model behaviour' in which they are likely to use the nearest available emergency exit or a more expected behaviour in which most users retrace their steps back towards the main entrance for the building. Figure 4 illustrates an application of the GES tool by analysing evacuation times when one of the emergency stairwells is blocked. The top line shows mean evacuation times under different occupancy levels when occupants are likely to retrace their route into the building. The lower line provides the same information for 'model' evacuations in which each occupant attempts to exit by the nearest available route. The difference between the 'model' and 'normal' mean evacuation times is much greater than for any other emergency stairwells. Hence, considerable efforts should be made to ensure that building occupants use this route rather than retracing their steps if they are to benefit from the timesavings indicated in Figure 4.

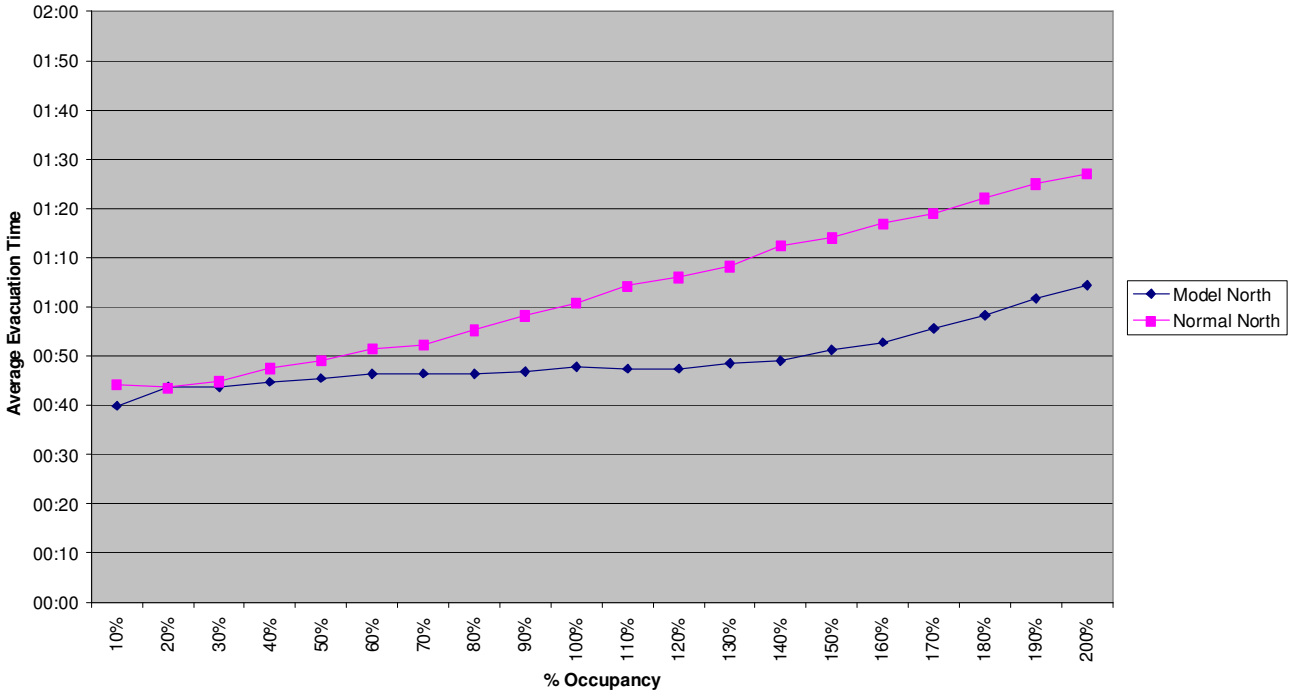
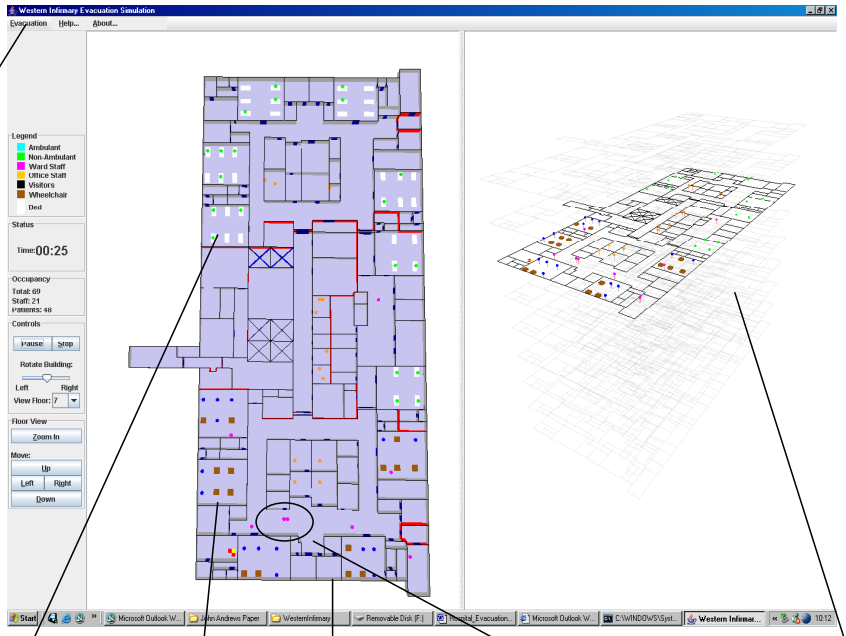


Figure 4: Graphing Mean Evacuation Times when the North Exit Route is Closed

Most existing simulation tools are tailored for the evacuation of large office blocks or entertainment complexes, including cinemas and sports stadiums. Others have been designed for trains, boats and airplanes. Some tools have been extended to support the simulated evacuation of healthcare institutions. For instance, Gwynne et al (2003) contrast the gathering of evacuation data and model development for a University and a Hospital Outpatient Facility. They argue that these two facilities 'employ relatively similar procedures: members of staff sweep areas to encourage individuals to evacuate'. However, the authors also identify numerous differences. Patients only began to leave once a member of the nursing staff instructed them to evacuate. Students were less dependent upon the actions of the staff. This study focused on outpatients. The differences in occupant behaviour between hospitals and other types of institution are more significant for simulations that consider in-patient care. For instance, flocking behaviours are often included in behavioural models for large public buildings. Occupants coalesce into larger groups and will tend to respond to an evacuation in similar ways. This emergent behaviour tends to be less of a feature in hospital evacuations where smaller numbers of patients and visitors may be directed to follow the horizontal evacuation procedures mentioned in previous paragraphs. Similarly, the models of individual behaviour are less important within this context. Individual assertiveness can be a significant factor when modelling the undirected response of individuals within an evacuation. However, it has far less of a role to play in hospital evacuations where staff have been trained to respond in a coordinated manner. Command hierarchies and roles are, typically, determined well before an evacuation through the preparation of detailed plans. They are reinforced through drills and exercises. In consequence, the development of hospital simulations must focus more on the modelling of plans and procedures than on the impact of individual assertiveness or on the emergent behaviours of large crowds. The GES, like most of the other tools mentioned above, was not specifically developed to simulate hospital evacuations. The following sections, therefore, describe the design and implementation of the Glasgow-Hospital Evacuation Simulator (G-HES) tool that is specifically intended to support the evacuation of hospital buildings.

3. The Glasgow-Hospital Evacuation Simulator (G-HES)

As mentioned, most simulators have been designed to model evacuations that are very different from the techniques that we have described for hospitals. There are some exceptions. For instance, Takenaka have developed the 'Assisted Evacuation Simulation System' (Jafari, Bakhadyrov and Maher, 2003). This is designed to simulate assisted evacuations across a range of environments and provides different occupant models for people are not capable of evacuating themselves. The tool enables users to vary the number of patients and helpers. It also simulates a range of evacuation methods including stretchers, wheelchairs and evacuation by helpers' supporting patients on both sides. Although this system provides sophisticated support for modeling the assisted evacuation of patients, it can be difficult to simulate some of the more detailed task allocations that are made in the complex evacuation plans of the case study. For example, one nurse is charged with exhaustively searching for the source of any alarm while colleagues use a whiteboard to coordinate other aspects of the evacuation. It is for this reason that the following pages describe the design and evaluation of an evacuation system that simulates a range of behaviors both for patients and staff.



Menu options enable users to alter the location of a hazard, such as a fire. Users can also alter the staffing levels available to move patients. Different proportions of ambulant, wheelchair and non-ambulant patients can also be specified. Evacuations can take up to an hour to complete if there are large numbers of non-ambulant patients and few staff so simulations can be set to run up to ten times 'faster than real time'.

White rectangles show beds, brown squares are wheelchairs.

Overhead plan of selected floor showing fire walls in red.

Purple squares show ward staff coordinating evacuation.

3D Floor plan with current floor giving aerial view of evacuation

Figure 5: The User Interface to the Glasgow-Hospital Evacuation Simulator (G-HES)

The project began by developing the 3D building model that is used in most evacuation simulators. This is especially important for hospitals where horizontal evacuation will lead to vertical evacuation when fires and other hazards jeopardize the safety of individual floors. Previous sections have mentioned that many public buildings now have electronic plans stemming from the increasing use of AutoCAD and similar products by architects. Tools, such as the GES, can semi-automatically read these during the construction of a simulation. Unfortunately, these plans are not always readily available for legacy buildings. They can also provide unreliable information given that the original infrastructure can be heavily modified as occupants remodel a building to support different activities. Semi-permanent structures and partition walls may not always appear of the plans that are supplied. For these reasons, the model illustrated in Figure 5

was developed by hand from paper plans that were then validated and cross-references through site visits that made use of digital photography for later off-site comparisons. We were particularly interested in the firewalls, illustrated in red on the previous image, because these denoted the boundaries for potential refuges where patients might be relocated during an emergency. We also had to model the difference between smoke resistant doors and doors that also provided protection against a spreading fire.

The initial stages in developing the G-HES involved analyzing the more general requirements that have been mentioned in previous sections. In particular, we conducted a number of focus groups with the Fire Officers and clinical staff who were to be the primary user group for the resulting application. Many of these discussions focused on the 'prototypes' that would be used to characterize the patients in each of the floors of the hospital. We began with four basic groups: 1) immobile patients who could not be moved from their beds; 2) Immobile patients who could be moved from their beds but only with considerable difficulty and an associated delay; 3) Immobile patients who could be moved with relative ease given the assistance of one or more members of staff; 4) Mobile patients able to move on their own with some staff directions. It can be difficult to predict precisely the distribution of patients within categories 1 to 4. Initial versions of the prototype simplified this taxonomy to consider ambulant and non-ambulant patients. However, future versions will return to these more elaborate distinctions. Similarly, it is possible to identify a number of categories within the nursing and clinical staff who are available to support an evacuation. The 'lead' nurse coordinates each evacuation. They will use a number of resources, such as a central whiteboard, to keep track of patient locations. The lead nurse can then dispatch their colleagues to initiate patient evacuation.

In addition to the more obvious occupant categories of patients and clinical staff the G-HES had to account for a number of other groups. For instance, many areas of the hospital are staffed by administrators and managers who would not normally be directly involved in the evacuation of patients. They would. However, receive annual training in evacuation and fire fighting procedures. They would also be familiar with the main emergency exits. However, as we have seen in the analysis of previous evacuations, it cannot be assumed that everyone in this category would choose to use these fire exits in preference to the main entrance routes into their areas within the hospital complex. As with all categories of staff, the level of administrative support varies considerably over the working day. Hence any simulation software must help its users differentiate between 'office hours' and other periods when less of these staff will be available.

There are significant numbers of visitors to some of the floors. However, these relatives and friends must, typically, restrict their visits to particular times. As development progressed, however, we quickly realized that the procedures and practice varies between different units. It is, therefore, possible for users of the simulator to specify how many visitors there will be on a particular floor prior to running the simulator. Similarly, it is possible to vary the occupants in floor that house out-patient's clinics by altering the distribution between mobile patients, who represent frequent visitors to the clinic, and visitors, who can be used to represent individuals who are new to the clinic and hence may not be familiar with the building layout.

As mentioned, there is an ordering that helps to determine evacuation priority. There is an expectation that office staff will require minimal supervision during an evacuation. All patients in immediate danger are moved first. Next ambulatory patients and visitors are moved. Wheelchair patients may be groups together and then moved gradually to a place of safety. Finally, non-ambulatory patients will be moved typically with moving those who can be transferred most easily before those who require significant additional preparation. The implicit objective at each stage is to maximize the number of people who can be moved to safety in the shortest available period of time. In addition to modeling these task priorities, it is important for the simulation to consider the timing delays associated with each of these evacuations. Firstly there is a preparation overhead in helping a patient to evacuate. Approximate timings are provided in Table 2. In computational terms, these delays are represented as probability distributions and Monte Carlo techniques can help to determine the real-time duration of any delay. These distributions can be assessed using experimental techniques. They can also be validated using a form of task analysis with staff focus groups given the difficulty of moving critically ill patients in a simulated exercise.

	Patient Category	Minimum delay (Seconds)	Maximum delay (Seconds)
1	Immobile patients who could not be moved from their beds (depending on associated instrumentation).	180	900
2	Immobile patients who could be moved from their beds but only with considerable difficulty and an associated delay (eg to a wheelchair)	180	900
3	Immobile patients who could be moved with relative ease given the assistance of one or more members of staff.	60	180
4	Mobile patients able to move on their own with some staff directions (accounting for telling them what is about to happen).	30	90

Table 2: Initial Preparation Times for Patient Evacuation

Once staff have initiated the evacuation of a patient, it is important for the simulator to determine their average walking speed. There have been many studies into average walking speeds during evacuations (Johnson, 2005). This work has, for example, looked at the manner in which we will slow down to accommodate different crowd densities. There has been relatively little research into the impact of walking speed on hospital evacuations. This creates several important problems. In particular, the relative age and physiological capacity of nursing staff is important given the problems of fatigue and of working in smoke filled environments performing tasks that involve considerable effort to complete. The initial simulations assumed a walking speed of between 2 and 0.04 meters per second. Again, Monte Carlo techniques can be used to assign particular speeds. Table 3 illustrates the results from a number of simple empirical tests to determine how these initial speeds should be modified depending on whether nursing staff were on their own or assisting in the movement of a wheelchair or a bed.

All timings are approximate for 10 meter distance.	Slow (seconds)	Medium (seconds)	Fast (seconds)
Nurse alone	16	12	8
Nurse with Wheelchair	20	16	12
Nurse with Bed	35	25	20

Table 3: Approximate Timings for Patient Evacuation over a Ten-Meter Distance

At present, the G-HES tools do not account for fatigue effects. However, the existing software could easily be enhanced to include a clock-based modifier to slow the speed of each nurse the longer that they participate in an evacuation. It is also important to emphasize the approximate nature of these timings. They depend upon the layout of the route being traveled. In this case we assumed that there were no obstacles and, in particular, the movement of the bed did not require any complex rotations to clear sharp corners. Similarly, the timings given above reflect the equipment available to staff on a particular floor of a particular hospital. The ease with which beds can be moved, in particular, depends on the particular model and degree of maintenance provided. For instance, the beds in our case study measured approximately 1 meter (38 inches) by 2.2 metres (86 inches). Wheelchairs were approximately 0.75 metres (30 inches) by 0.75 metres (30 inches). However, there were several different models. Some wheelchairs were heavily upholstered and more similar to a moveable armchair. Others were based around more conventional metal frames. Initial observations showed considerable variation both in the time to move patients between beds and the wheel chairs and to negotiate potential obstacles under ideal conditions; without smoke etc.

One of the most difficult problems for any simulation is to determine how human behavior will change over time as events unfold during an evacuation. In large group systems, such as the Boyd Orr auditorium system illustrated in Figure 3, individuals alter their behavior in response to changes in direction and speed

within the crowd. It is for this reason that the GES uses Monte Carlo techniques where the likelihood that an individual will move in a particular interval is determined amongst other things by the speed and proximity of their neighbors. Such issues are less important in the simulation of hospital evacuations; crowds are less likely to occur except for bottlenecks close to common evacuation routes during visiting hours or in outpatient clinics. In contrast, it is important to account for the ways in which nursing staff will alter their response to an emergency within the constraints provided by 'horizontal evacuation' procedures and related hospital policies. For instance, many healthcare institutions are deliberately designed around a grid-structure where wards and rooms can be accessed from two different directions along common corridors. Nursing staff, therefore often have to choose between several alternate routes between a patient's room and a place of safety. Any simulation should account for those factors that are likely to influence the nurses' decision to use a particular corridor. For example, they should not normally lead patients along corridors that pass the seat of a fire. It must also account for the occasional situations when nursing staff select a more dangerous or slower route, either because they lack critical information or because they make a mistake. Additional complexity is introduced by a requirement that staff should continue to make intelligent decisions about where to move patients as a fire progresses and more routes become blocked.

The implementation of the nursing staff that drives the evacuation of the hospital is based around autonomous threads. The program creates an independent process for each individual. These processes can communicate through a form of message passing; the 'actions' that each nurse performs are implemented based on the represented state of the environment. A form of reactive route finding is implemented for each nurse using the A* algorithm that was first developed within the field of Artificial Intelligence. This assumes that the simulated nurse can identify each of the possible moves that they can make from their current location. They rank each of these moves and then only go on to consider the next set of available moves from the top ranked adjacent position. In this way, their planned route gradually grows as they always pick the best next step for further consideration. If a potential route becomes blocked then it may be necessary to consider the second route in the list of preferences. The success of the algorithm depends upon the choice of an appropriate heuristic. Euclidian distance can be used. Alternatively, more detailed information about the layout of the hospital can also be used to guide the evacuation movements. Recall that an independent thread represents each nurse. Each nurse will also be employing his or her own independent navigation strategy. It is, therefore, possible that contention will occur if, for example, two nurses attempt to move two beds along the same narrow corridor. This is entirely to be expected and specialist negotiation algorithms must then be used to resolve the bottleneck that is also a feature of 'live' evacuation drills. Brevity prevents a full introduction to the range of programming techniques that were used and the interested reader is directed to (Ashraf et al, 2003).

Figure 6 illustrates two key features of the hospital evacuation simulator. The image on the left shows a single panel from the G-HES configuration manager. Users can either alter the total number of staff and patients in different categories or they can alter a ratio of the current maximum occupancy and staffing levels. This interface can also be used to determine the anticipated number of people in the building for simulation runs at particular times of day. The other options available through the tabs on the left-hand image help the user to control the location of the fire. The 'General' option controls the speed of the simulation and allows a certain degree of lower level control over the procedures and route finding algorithms employed by the staff during an evacuation. In contrast, the image on the right of Figure 6 illustrates the output from a single run of the simulation. As can be seen, this run took a total of 17 minutes and 23 seconds to move all of the patients to a place of safety. This illustrates the importance of the option to run simulations at up to ten times their normal speed in order to assess a range of different non-deterministic evacuation behaviors in a particular configuration. The termination of an evacuation run in the context of a hospital evacuation raises a number of questions that do not arise in more conventional simulators. For example, in an auditorium or office block a run can be terminated when all of the occupants have safely exited from a building. In a hospital, however, this is not the case. Horizontal evacuation techniques rely upon the movement of patients to compartments that have a safe exit and that are protected by fire resistant walls and doors. It follows that the safety of patients and staff can be undermined even when this has been achieved. A fire or other hazard may gradually spread into areas that are immediately adjacent to this temporary place of safety. The users of the hospital evacuation simulator, therefore, have the option to restart an evacuation with the fire located in a different position in the

building. Staff must then move their patients again. In practice, however, there is an assumption that emergency help will have arrived before such a 'last resort' action would ever be needed.

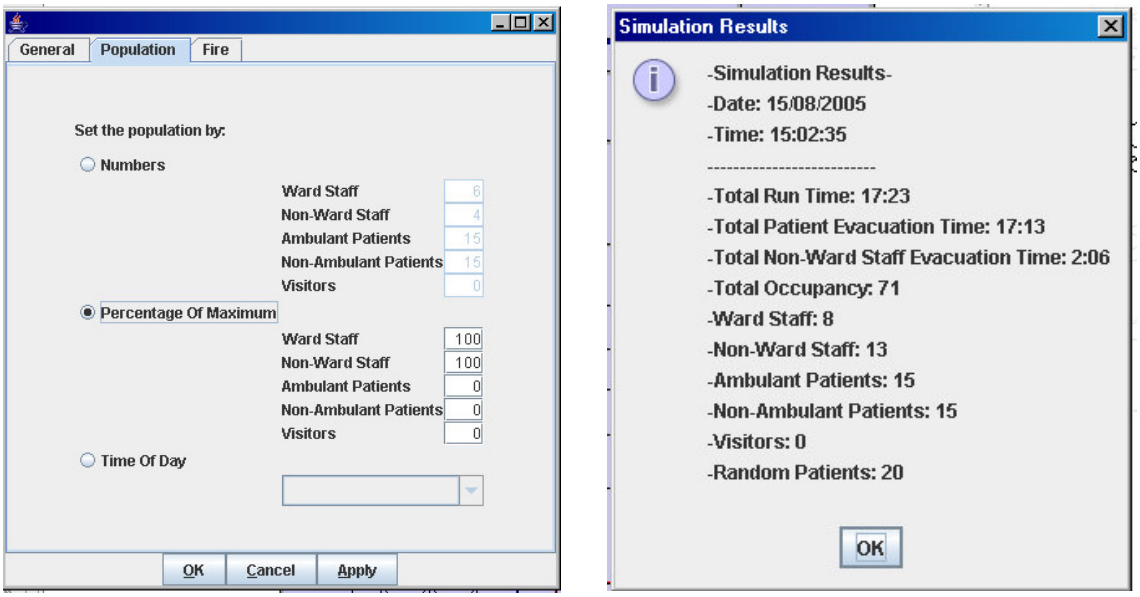


Figure 6: Option Panel and Results Dialogue from the Glasgow-Hospital Evacuation Simulator

The resulting simulator can be used in a range of ways. One immediate application was to explore what might happen to evacuation times with different profiles of ambulant and non-ambulant patients under the given staffing regime within particular areas of the hospital. As mentioned previously, simulator allows for non-determinism both in the patient profile and in the concurrent interaction between staff as they plan the best evacuation routes for a particular hazard. We, therefore, began to apply the tool by examining ten separate runs for the current staffing level of six nurses faced with different proportions of ambulant and non-ambulant patients. The results are shown in table 4.

Number of Non-Ambulant Patients	**Number of Ambulant Patients**	**Mean Evacuation time in seconds (Min:Sec)**	**Standard Deviation in seconds (Min:Sec)**
30	0	2643 (44:03)	257 (4:17)
25	5	1749 (29:09)	205 (3:25)
20	10	1439 (23:59)	189 (3:09)
15	15	1105 (18:25)	86 (1:26)
10	20	801 (13:21)	75 (1:15)
5	25	707 (11:47)	64 (1:04)
0	30	470 (7:50)	54 (0:54)

Table 4: Evacuation Times for Day Staff of 6 Nurses with 10 Runs for Each Patient Distribution

Table 5 continues the analysis showing the same means and standard deviations for different combinations of ambulant and non-ambulant patients. In contrast to Table 4, this illustrates the increased evacuation times associated with the reduced staffing levels that typically hold at night. It should be stressed that these figures are illustrative. As mentioned previously, agency staff are used more frequently to fill these shifts. The simulations do not currently take into account any additional overheads associated with reduced levels of staff training. Similarly, they do not consider the additional complexity of rousing ambulant patients from sleep when they may be under additional sedation. Finally as mentioned previously, we do not explicitly take into account the additional fatigue that may be expected if a small number of staff are involved in an evacuations that would require almost an hour to complete. The programming of these

additional factors would be relatively straightforward compared to the synchronization techniques needed to implement nursing staff as a parallel processes. There are several reasons why these factors have not been explicitly modeled. Unlike the figures for daylight evacuations it is far harder to conduct nighttime validation exercises through live drills. It is unclear whether it would ever be possible or ethical to obtain staff participation to assess fatigue in an exercise involving non-ambulant patients where simulation results indicate it might take an hour or more.

Number of Non-Ambulant Patients	Number of Ambulant Patients	Mean Evacuation Time in seconds (Min:Sec)	Standard Deviation in seconds (Min:Sec)
30	0	3445 (57:25)	363 (6:03)
25	5	2976 (49:36)	279 (4:39)
20	10	2703 (45:03)	253 (4:13)
15	15	2357 (39:17)	234 (3:54)
10	20	1991 (33:11)	226 (3:46)
5	25	1723 (28:43)	244 (4:04)
0	30	1343 (22:23)	227 (3:47)

Table 5: Evacuation Times for Night Staff of 3 Nurses with 10 Runs for Each Patient Distribution

In spite of the caveats raised in the previous paragraphs, the results from the evacuation simulator provided important information to hospital administrators and managers as they assessed the risks associated with current staffing levels given different combinations of ambulant and non-ambulant patients. Given the difficulties of conducting 'live' drills and validation exercises, the greatest contribution of this type of tool need not lie in the accurate prediction of evacuation times as an outcome in itself. In contrast, our experience has shown that it can provide the greatest benefits in promoting a risk-based approach to the planning of evacuation exercises. Our preliminary figures for the night-time evacuation showed that there was an urgent need to determine whether current wards with, for example a mix of 10 ambulant to 20 non-ambulant patients, could be evacuated safely given the range of hazard scenarios considered in the emergency evacuation plans. Hence the use of the simulator drove another round of risk assessment that included the need to run night-time live 'drills' to validate the initial findings.

4. Conclusions and Further Work

The safety of large public buildings has become a pressing concern following recent and terrorist actions in Europe and the United States. This has led many regulatory and governmental agencies to advocate a risk-based approach to evacuation. The owners and operators of these buildings must demonstrate that they have taken actions to mitigate the most serious hazards that could prevent a successful evacuation. Unfortunately, it can be difficult to apply existing risk assessment techniques in this domain. Fault trees and FMECA can be used to represent and reason about potential problems. However, it is hard to assess the criticality or even the consequences of hazards, such as a fire exit becoming blocked or of a fire occurring during times of day with an increased occupancy or reduced staffing level. Some of these problems stem from the difficulty of conducting a program of 'live' evacuation exercises. Many buildings are now occupied by thousands of staff. Evacuation drills can endanger those occupants with pre-existing cardio-vascular conditions. They can also prove to be particularly disruptive to the financial and healthcare industries that must provide 24/7 support to their clients.

This paper has described how simulation software can be integrated into a risk-based approach to the evacuation of large public buildings. These tools can be programmed with models that are informed by an analysis of evacuation procedures and also be observations of human behaviour during both 'real' evacuations as well as drills. For example, timings taken from an evacuation exercise can be used to 'fine tune' the predictions made by the simulator. This is an iterative process because the results from a simulation can then also be used to focus subsequent 'live' evacuation exercises. The results of this process can then provide evidence for risk assessments that are structured using more conventional techniques, such as Fault Tree Analysis. The likelihood of particular combinations of hazardous events can

be demonstrated by reference to previous accident reports and to live exercises. Where this information is partial or cannot ethically be obtained then computer-based simulations can be used.

Although this 'risk-based' approach to evacuation does not seem to have been explicitly written-up before, it shares much in common with the use of simulation in other engineering disciplines. We have, therefore, chosen to apply the technique in an innovative way by developing the Glasgow-Hospital Evacuation Simulator (G-HES) that is explicitly intended to model the evacuation of a large hospital building. This decision justified by the ethical and practical problems associated with 'live' exercises involving patients. These institutions also pose a considerable risk in terms of the relatively high frequency of fires and also the high potential consequences illustrated by several recent accidents. In many ways, these buildings pose extreme challenges. Occupant models must reflect the complex movement strategies that are devised to ensure that as many patients are moved as quickly as possible to a place of safety. The simulations must also consider the behaviour of other occupants including visitors and administrative staff. Finally, it is important to consider the impact of ambulant and non-ambulant patients where staff may be forced to first prepare patients to be evacuated and then move them using beds and wheelchairs.

We have implemented the G-HES using concurrent programming techniques to model nursing staff as they implement a horizontal evacuation strategy. This technique has been combined with independent route finding algorithms so that staff will automatically alter their actions to 'work around' their colleagues' activities. It is still possible, however, for contention to occur if colleagues try to move several patients along the same corridor. These algorithms also account for changes in strategy as fires spread to block previous evacuation routes. Again, however, the non-determinism in the application can capture periodic mistakes in which staff expose both themselves and patients to unnecessary risks, for instance, by moving down corridors that had previously been safe to navigate. G-HES can also be extended to use Monte Carlo techniques to determine the precise delays that are incurred as staff prepare patients to be moved and then move them away from a hazard. The rate of movement is non-deterministically assessed using speed distributions obtained by empirical studies of staff in the case study institution. Finally, the completed simulator has been applied to assess the amount of time that would be required to evacuate a mixed profile of ambulant and non-ambulant patients given the typical staffing levels both on day shifts and during the night. The results of this study illustrate the need for an iterative approach by motivating further 'live' evacuation drills to confirm the predicted results for nighttime evacuations. The insights obtained from the simulation proved to be crucial in justifying drills that might otherwise have been dismissed as unjustified given the ethical concerns over such exercises. Hence the simulators not only support a risk-based approach to evacuation planning, they also help to inform a risk-based approach to the planning of evacuation exercises.

References

F. Ashraf, J. Johnston, C. McAdam, G. Mckinlay and M. Wilson, The Hospital Evacuation Simulator, Technical Report, Department of Computing Science, University of Glasgow, June 2005.

UK Atomic Energy Authority, A Technical Summary of the AEA Egress Code, technical report AET/NOIL/27812001/002(2), Issue 1, Warrington UK, 2002.

J.L. Bryan, Human behavior in the MGM Grand Hotel fire. Fire Journal. 76:37–48, March 1982.

UK Building Research Establishment, Evacuation Modeling: GridFlow and CRISP, Watford, UK, 2004. Available on: http://www.bre.co.uk/frs/

S. Chamberlain, M. Modarres and F. Mowrer, Compressed Natural Gas Bus Safety: A Qualitative and Quantitative Risk Assessment, Technical Report, Center for Technology Risk Studies, University of Maryland, USA, CTRS-MC1-02, May 2002. Available on http://www.enre.umd.edu/ctrs/report.pdf, August 2005.

P. Edelman, E. Herz and L. Bickman, A model of behavior in fires applied to a nursing home fire. In: Canter D, ed. Fires and fire behavior. John Wiley and Sons, Chichester, UK, 1980.

S. Gwynne, E.R. Galea, J. Parke and J. Hickson, The Collection of Pre-Evacuation Times from Evacuation Trials Involving a Hospital Outpatient Area and a University Library Facility. In D. Evans (ed.) Proceedings of 7th IAFS Symposium WPI, USA, pp877-888, 2003.

International Association of Fire Chiefs, Healthcare Fire Safety Round Table Forum. Washington, DC, USA. 2004. Available on:

http://www.fire.gov/newsletter/winter2005/IAFC_RdTable/healthcare.pdf, August 2005.
M. Jafari, I. Bakhadyrov and A. Maher, Technological Advances in Evacuation Planning and Emergency Management: Current State of the Art, Technical Report to the Federal Highway Administration, EVAC-RU4474, Center for Advanced Infrastructure & Transportation (CAIT), Rutgers University, New Jersey, USA, March 2003. Accessed on http://www.cait.rutgers.edu/finalreports/EVAC-RU4474.pdf, August 2005.
C.W. Johnson, A Handbook of Accident and Incident Reporting, Glasgow University Press, Glasgow, 2003. Available on-line at: http://www.dcs.gla.ac.uk/~johnson/book
C.W. Johnson, Applying the Lessons of the Attack on the World Trade Center, 11th September 2001, to the Design and Use of Interactive Evacuation Simulations, In Proceedings of ACM CHI 2005, ACM Press, New York, USA, 651-660, 2005.
N. Latman, TCPP Personality Profile, In The Fourth Triennial International Fire and Cabin Safety Research Conference, 15-18 November 2004, Parque das Naçoes Conference Centre, Lisbon, Portugal, 2004.
P.M. McCarthy and K.A. Gaucher, Fire in the Operating Room—Developing a Fire Safety Plan, AORN Journal (Journal of the Association of Perioperative Registered Nurses), 79:3, March 2004.
H. Muir, Research into the factors influencing survival in aircraft accidents. The Aeronautical Journal, May 1996, 177-181.
M Owen, E Galea and P Lawrence, The EXODUS Evacuation Model Applied to Building Evacuation Scenarios. Journal of Fire Protection Engineering 1996, Vol.8(2), pp 65-86.
G. Proulx, Occupant Behaviour and Evacuation. Proceedings of the 9th International Fire Protection Symposium, Munich, 25-26, pp 219-232, 2001.
P.A. Thompson and E.W. Marchant, Computer and fluid modelling of evacuation. *Safety Sci* **18**, pp. 277–289, 1995.
D. Tong and D. Canter, The Decision to Evacuate: A study of Motivations which Contribute to Evacuation in the Event of Fire. Fire Safety Journal 9:257-265. 1985.
US Joint Commission on Accreditation of Healthcare Organizations, Preventing Surgical Fires, Sentinel Event Alert, Issue 29, 24th June 2003.
Accessed on http://www.jcaho.org/about+us/news+letters/sentinel+event+alert/sea_29.htm, August 2005.
US National Conference of States on Building Codes and Standards (NCSBCS), News release: At NCSBCS Annual Conference, Construction Industry, Building Regulators Discuss Coordinated Actions Critical to Economic Competitiveness, Public Safety and Homeland Security, 2002.
Available on: http://www.ncsbcs.com/
US National Fire Protection Association (NFPA), Summary Fire Investigation Report—Hospital Fire, Brooklyn New York, 1/9/1993, Department Secretary for Fire Investigation, National Fire Protection Association (NFPA),
Accessed on http://www.mdsr.ecri.org/summary/detail.aspx?doc_id=8164, August 2005.
US National Fire Protection Association (NFPA), Summary Fire Investigation Report---Hospital Fire, Petersburg, Virginia 5 People Killed, 12/31/1994, Department Secretary for Fire Investigation, National Fire Protection Association (NFPA).
Accessed on http://www.writer-tech.com/pages/summaries/summpetersburg.htm, August 2005.
US National Fire Protection Association (NFPA), NFPA 1600: The National Preparedness Standard - NFPA 1600 gives businesses and other organizations a foundation document to protect their employees and customers in the event of a terrorist attack. National Fire Protection Association Journal, January/February 2005. Accessed on http://www.nfpa.org, August 2005.
US National Fire Protection Association (NFPA), Technical Committee on Fire Risk Assessment Methods, Guide for the Evaluation of Fire Risk Assessments, 2004 edition, N FPA 551, Accessed on http://www.nfpa.org, August 2005a.
US Occupational Safety and Health Administration, OSHA Factsheet: Evacuating High-rise Buildings, US Department of Labor, Washington DC, 2003. Available on http://www.osha.gov
UK NHS, Fire Safety Policy, Directorate of Finance, Health Department, Scottish Executive NHS HDL(2001) 20, 2001. Available on http://www.show.scot.nhs.uk/sehd/mels/HDL2001_20.pdf, August 2005.

Root-Cause Analysis for Complex Security Incidents

C.W. Johnson

Dept. of Computing Science,
University of Glasgow,
Glasgow, Scotland.
http://www.dcs.gla.ac.uk/~johnson
johnson@dcs.gla.ac.uk

ABSTRACT
There is an increasing need for incident response to look beyond the immediate causes of security violations. For example, the US Department for Homeland Security has commissioned a number of recent reports into the 'root causes' of adverse events ranging from denial of critical infrastructure to barriers for security information transfer between Federal agencies. The US Department of Energy has also established the Information Security Resource Center to coordinate the 'root cause analysis' of security incidents. A recent report by the Harvard Business School (Austin and Darby 2003) highlighted several commercial initiatives to understand not simply what went wrong in any single previous incident but also to identify any further underlying vulnerability. A common theme in all of these initiatives is to go beyond the specific events of a particular security incident and to identify the underlying 'systemic' technical, managerial and organizational precursors. Unfortunately, there are relatively few established tools and techniques to support the 'root cause' analysis of such incidents. This paper, therefore, provides an introduction to V^2 (Violation and Vulnerability) diagrams. The key components of this technique are deliberately very simple; the intention is to minimize the time taken to learn how to exploit this approach. A complex case study is presented. The intention is to provide a sustained analysis of Rusnak's fraudulent transactions involving the Allfirst bank. This case study is appropriate because it included failures in the underlying audit and control mechanisms. It also stemmed from individual violations, including the generation of bogus options. There were also tertiary failures in terms of the investigatory processes that might have uncovered the fraud long before Allfirst and AIB personnel eventually detected it.

Keywords: Root-cause analysis; Security violations; Accident analysis.

INTRODUCTION

It seems unlikely that we will ever be able to eliminate security related incidents across a broad range of public and private organizations. The continual pressures for additional functionality through technological innovation create vulnerabilities that can be difficult to anticipate or guard against. The US military describe how during Operation Desert Storm and Desert Shield, 'perpetrators who were thousands of miles away illegally accessed dozens of U.S. military systems… sophisticated break-in techniques were employed to obtain data about U.S. troop movements, ordnance systems, and logistics…new security vulnerabilities that expose systems and networks to unauthorized access and/or deny service are constantly being discovered' (Dahlgren, 2002). Given that it is impossible to achieve total security, it is important that organizations plan their response to those attacks that do occur. For instance, the CISCO (2003) 'Best Practices White Paper' on network security urges companies to collect and maintain data during security incidents. This information can be used to determine the extent to which systems have been compromised by a security attack. It can also be critical to any subsequent legal actions; "if you're interested in taking legal action, have your legal department review the procedures for gathering evidence and involvement of the authorities. Such a review increases the effectiveness of the evidence in legal proceedings". These recommendations reflect the current 'state of the art' in incident investigations. The focus is on the groups and individuals who perpetrate an attack rather than the underlying technical, managerial and organizational factors that create 'systematic' vulnerabilities in complex systems.

There is a growing realization that security investigations must examine the root causes of security incidents. A number of organizations already recognize the importance of this 'lessons learned' approach to security incidents. For example, the Los Alamos National Laboratory adopted this approach in the aftermath of a series of security related incidents involving information about nuclear weapons research. The mishandling' of two computer hard drives containing classified information led the director of the laboratory to report to the Senate Armed Services Committee. This report focused on the individual human failures that were identified as root causes. However, it also consider the contributing factors that included the 'government-wide de-emphasis on formal accounting of classified material that began in the early 1990s, which weakened security practices and created an atmosphere that led to less rigor and formality in handling classified material'(Roark, 2000). These and similar findings have led the US government to focus more directly on the different factors that contribute to the underlying causes of security vulnerabilities. The Government Security Reform Act (2001) transferred the Federal Computer Incident Response Capability (FedCIRC) from the National Institute for Standards and Technology (NIST) to the General Services Administration (GSA). As part of this move, the GSA was charged to identify patterns in the causes of security incidents (Lew, 2001).

Similar trends can be observed in commercial organizations, especially business consultancies. For instance, Price Waterhouse Cooper (Skalak, 2003) recently issued a brief on understanding the root causes of financial fraud. They argued that 'the key for companies is to use a global risk paradigm that considers the root causes of financial fraud, corporate improprieties and potential regulatory malfeasance arising from different markets, and therefore different risk environments, in which global enterprises operate'. Although their focus is on the wider aspects of fraud and not simply of security, the Investigations and Forensic Services group within PWC have argued that a wider form of 'root cause' analysis represents a new paradigm for the investigation of security incidents. The intention is to probe beyond the specific violations of external agencies and junior staff members to look at the wider organizational problems that created the context and opportunities for these threats to be realized. Several accountancy firms in the US and Europe have adopted a similar perspective as they begin to examine the consequences of recent corporate scandals. In particular, they have looked beyond the individual (mal-)practices in particular cases. It has been argued that 'controls, no matter how sound, can never prevent or completely limit persons in high places from circumventing controls or prevent or detect all fraud …auditors do not guarantee discovery of all fraud but provide only reasonable assurance of the absence of material fraud…there have been too many instances of fraud, transactions in excess of authorized limits, and other negative events while controls were thought to be in place or auditors present to permit acceptance of these contentions. Many factors have created the current quandary. They require clear understanding and careful response for auditors and organizations they serve to rebuild the level of public confidence previously enjoyed' (Rabinowitz, 1996).

PRIMARY, SECONDARY AND TERTIARY FACTORS IN SECURITY INCIDENTS

The previous quotations argue that specific violations that lead to security incidents often form part of a more complex landscape of external threats, managerial and regulatory failure, of poor technical design and of operational inadequacies. Mackie (1993) uses the term 'causal complex' to describe this causal landscape. Although he was looking purely at the philosophy of causation, it is possible to apply his ideas to clarify some of the issues that complicate the investigation of security incidents. Each individual factor in a causal complex may be necessary for an incident to occur but an attack may only be successful if they happen in combination. Several different causal complexes can lead to the same outcomes even though only one may actually have caused a particular incident. For instance, unauthorized trading might not be detected because of insufficient oversight, collusion or through oversight that was ineffective. It is for this reason that most security investigations consider alternate scenarios in order to learn as much as possible about the potential for future failures. In our example, an investigation might look at the potential impact of collusion even if a particular incident stemmed from inefficient oversight. These high-level arguments are grounded in Microsoft (2003) technical advice for security audits: "During security risk identification, it is not uncommon for the same condition to have multiple consequences associated with it. However, the reverse also may be true there may be several conditions that all produce the same consequence. Sometimes the consequence of a security risk identified in one area of the organization may become a risk condition in another. These situations should be recorded so that appropriate decisions can be made during security risk analysis and planning to take into account dependencies and relationships between the security risks".

Mackie goes on to argue that we often make subjective decisions about those factors that we focus on within a causal complex. The term 'causal field' refers to those factors that an investigator considers relevant to a particular investigation. If a cause does not appear within this subjective frame of reference then it is unlikely that it will be identified. This philosophical work has empirical support from the findings of West-Brown et al's (2003) study into the performance and composition of Computer Security Incident Response teams. They describe the difficulties of ensuring that organizations and individuals broaden their view of the causal field to identify the different vulnerabilities that are exposed in the aftermath of security incidents. The problems of determining alternate causal fields are exacerbated by a number of factors identified by Meissner and Kassin (2002). They show that rather than improving accuracy in detecting deceit, training and prior experience make individuals more likely to identify 'deceit' rather than 'truth' in laboratory conditions. In other words, investigators cannot easily be trained to accurately identify whether evidence about the causes of an incident is true or not. Previous experience simply increases the likelihood that they will doubt the veracity of the information they obtain.

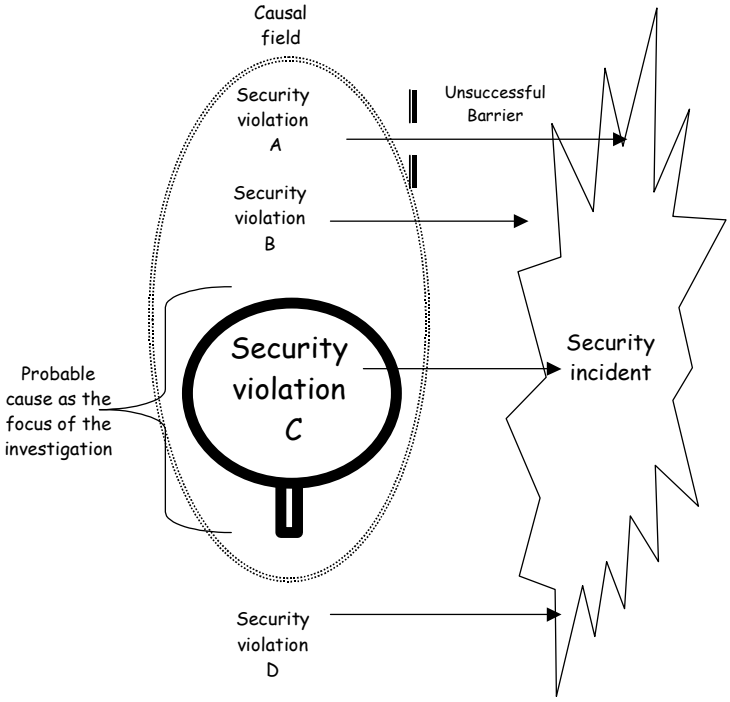


Figure 1: Causal Fields and Primary Security Violations

Figure 1 provides an overview of Mackie's ideas. The causal field in this case concentrates on violations A, B and C. The term 'violation' refers to any act or omission that contravenes security requirements within an organization. Within the causal field, we can focus on particular issues that we raise to the status of 'probable causes'. This is illustrated by the magnifying glass. For example, an investigator might be predisposed to look at the relationship between front office traders and back-office settlement staff. This would be illustrated by the focus on the potential primary violation C in Figure 1. However, the causal field may not encompass a sufficient set of conditions and in this case Primary violation D is not within the range of issues being considered by the investigator. For instance, if the investigation focuses on the manner in which a rogue trader exploited vulnerabilities in reporting systems then correspondingly less attention may be paid to the role of other team members in detecting potential losses.

It is important to emphasize that this broader view of causation does not absolve individuals from responsibility for their role in security incidents. It is, however, important to recognize the diversity of other features within the causal complex of security incidents. In particular, the opportunities for individual violations are typically created by organizational and managerial problems. Individual criminal acts often form part of a more complex series of causes that are collectively sufficient for an incident to occur (Reason, 1997). In other words, many failures stem from 'second order' vulnerabilities. These describe problems that do not directly cause an adverse event but can help to create the conditions in which a security incident is more likely to occur.

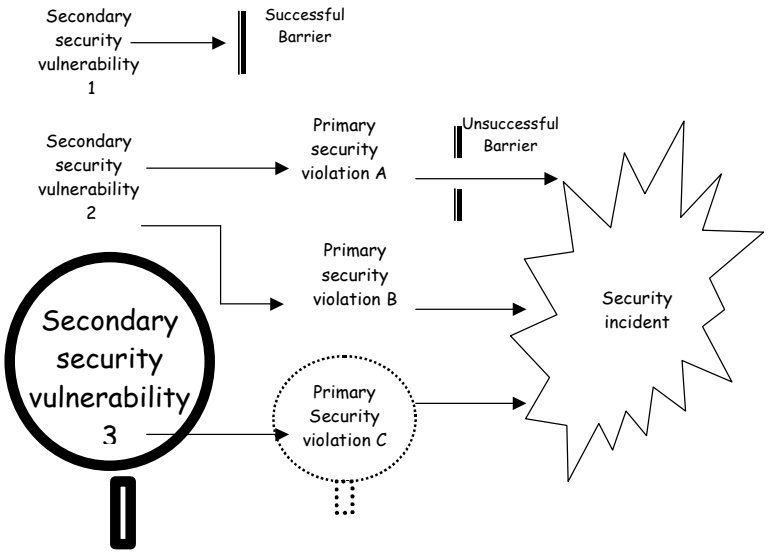


Figure 2: Causal Fields and Secondary Security Vulnerabilities

Figure 2 provides an overview of secondary security violations. As can be seen, these problems contribute to primary failures. As we shall see, a lack of oversight in the separation between front office traders and the back-office staff responsible for settling accounts, represented by secondary failure 2, can create the vulnerabilities that are exploited by a rogue trader. This is illustrated by primary violation B in Figure 2. Alternatively, internal inspections by compliance teams following the model recommended by the Bank of England after the Baring collapse might help to detect such secondary vulnerabilities before they can be exploited. The successful barrier to secondary violation 1 in Figure 2 would illustrate this. An important aim of this paper is to extend the causal field of security investigations to consider these secondary causes of adverse events. This is illustrated in Figure 2 by moving the magnifying glass to the left. The dotted ellipse used to denote the causal field in Figure 1 could also be redrawn to show the extended scope of an investigation in this figure. Our emphasis on secondary violations is intended to guide the composition of a causal field, which Mackie argues can be a subjective and arbitrary process. These underlying secondary organizational, managerial and regulatory issues are an increasingly common factor in the assorted lists of 'contributory factors' that appear in security incident reports. We would, therefore, argue that these secondary violations deserve greater and more sustained attention.

To summarize, first order security violations lead directly to an incident. They are cited as the probable cause when, for instance, an individual attempts to place an unauthorized transaction. In contrast, secondary security vulnerabilities make these primary actions more likely. For example, inadequate management supervision can increase a rogue trader's perception that their actions will not be detected. The increasing prominence of these secondary factors in regulatory reports suggests that more attention should be played to their role in the causal fields that guide security investigations.

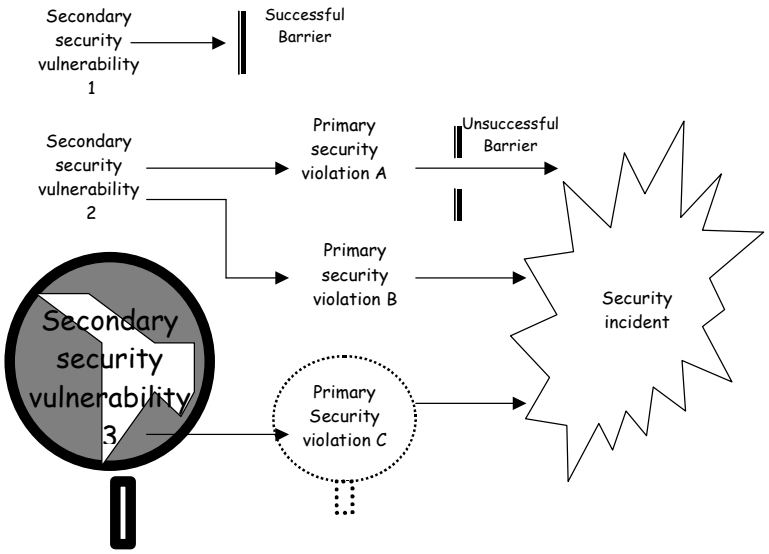


Figure 3: Causal Fields and Tertiary Investigatory Failures

The broken lens of the magnifying glass in Figure 3 illustrates a final form of failure that complicates the analysis of security incidents. Tertiary failures complicate the investigators' use of logs and other forms of evidence to reconstruct the events leading to a security incident. These problems need not directly lead to an incident nor do they make an incident more likely. However, inadequate investigatory procedures and tools can make it far less likely that investigators will consider an adequate range of factors within the causal complex of a security incident. In consequence, any subsequent analysis may overlook some vulnerabilities and violations. The following pages, therefore, present techniques that investigators can use to avoid these tertiary problems when they seek to identify the primary and secondary causes of security incidents.

THE CAUSAL ANALYSIS OF SECURITY INCIDENTS

It is clearly important that we learn as much as possible from those incidents that do take place if we are to reduce the likelihood and mitigate the consequences of security violations. A number of different tools and techniques can be used to support the analysis of these incidents. For instance, Julisch (2003) summarizes research into automated intrusion detection. He argues that over 90% of all alarms can be attributed to just over a dozen root causes. In consequence, rather than responding to individual alarms, investigators should focus on these more generic root causes using clustering methods that support the human analyst in identifying the underlying factors behind these warnings. Although this approach provides means of automatically clustering certain aspects of previous incidents, it cannot easily be applied to identify patterns in the organizational and managerial precursors to adverse events. In particular, it can be difficult to identify appropriate ways for representing and reasoning about these factors in security related incidents. Stephenson's (2003) recent work on Colored Petri Nets for the analysis of 'digital evidence' avoids some of these limitations. He assessed the impact of the SQLSlammer worm on a multinational company. He was able to work back from the technical properties of the attack to identify the company's business processes that made them vulnerable to this security threat. The formal Petri Net notation provided a common language for representing and reasoning about these different levels of analysis and hence could be used to move from the specifics of this incident to more general root causes. However, this work is based on a modeling language that was originally developed to support the design of concurrent systems. In consequence, it provides little direct support for the identification of root causes and contributory factors. The use of this approach is almost entirely dependent on the skill and expertise of the analyst. The lack of any supporting analytical methodology for the analysis of security incidents also makes it likely that two investigators will reach very different conclusions about the causes of an individual incident. This can help to identify a range of issues in the aftermath of an adverse event. Such inconsistency can also help to undermine the conclusions and recommendations that are drawn from an investigation.

Kilcrece et al's (2003) work on organizational structures for security response teams reinforces the comments of the previous paragraph. It also highlights the consequences of the lack of methodological support for investigatory agencies. They argue "different members of the security team may conduct very different types of analysis, since there is no standard methodology". In consequence, it is likely that effort will be duplicated both within response teams and across organizations as they address similar types of incidents. The lack of coordination and agreed procedures for the dissemination of root cause analysis makes it likely that similar patterns of failure will not be detected. This suggests that vulnerabilities will persist even though individual violations are identified. Without sharing this causal and contextual information, Kilcrece et al argue that the longer term recovery process will take longer and cost more, "problems that could have been prevented will instead spread across the enterprise, causing more down time, loss of productivity, and damage to the infrastructure".

The US Department of Energy has recognized the importance of adopting appropriate methodologies for the root cause analysis of security incidents, particularly involving nuclear installations. OE Order 470.1 requires that this form of analysis be conducted and documented as part of any process "to correct safeguards and security problems found by Department of Energy's oversight activities" (Jones, 2000). The intention is to ensure that any vulnerabilities are corrected in an 'economic' and 'efficient' manner. These methods are documented in the Department of Energy's (2003) standard DOE-STD-1171-2003, the Safeguards and Security Functional Area Standard for DOE Defense Nuclear Facilities Technical Personnel. This requires that security personnel must demonstrate a working knowledge of root cause analysis techniques that can be applied to 'determine the potential cause of problems'. They must be able to explain the application of root cause analysis techniques. In particular, they must be familiar with a number of specific approaches including causal factor analysis, change analysis, barrier analysis as well as management oversight and risk tree analysis. More detailed technical coverage of the application of these approaches is provided by the DOE (1992) standard DOE-NE-STD-1004-92, Guidelines for Root Cause Analysis. The adoption of root cause analysis does not, however, provide a panacea. A recent US General Accounting Office (GAO) report observed, "despite their importance, these assessments and analyses have not always been conducted". The GAO argued that steps must be taken to ensure that Department of Energy staff and sub-contractors follow the recommended root cause analysis techniques in the aftermath of security incidents (Jones, 2000). In particular, it is important that staff be provided with sufficient training and case studies to enable them to apply techniques such as those described in the standard 1004-92.

The work of the US Department of Energy in the development of root cause analysis techniques has not been mirrored by similar developments in commercial and financial organizations. Recent interest in causal analysis from security consultancies, such as Price Waterhouse Coopers, and by regulatory organizations, including the Bank of England, has not led to any consensus about how such analysis should be performed. There is, therefore, a need to identify appropriate methodologies to probe beyond specific violations to identify the underlying 'secondary' vulnerabilities that create the context for most security incidents. It is for this reason that the following paragraphs present a case study in the application of root cause analysis techniques to a large-scale fraud investigation. The aim is to determine whether the tools and methods that have been developed by the US Department of Energy for investigations into nuclear security incidents might be more widely applied within the commercial sector. Later sections will motivate the decision to use these particular techniques. For now it is sufficient to observe that accident and incident analysis within the field of safety-critical systems have been supported by a vast range causal investigation tools. Many of these are summarized in Johnson (2003). In contrast, we have chosen to focus on those approved by the US DOE because these techniques are well documented and have at least a limited track-record within the limited field of nuclear security investigations.

OVERVIEW OF THE ALLFIRST CURRENCY TRADING LOSSES

The remainder of this paper is illustrated by a case study involving the loss of approximately $750 million in currency transactions from Allfirst, a subsidiary of Allied Irish Bank. This case study is appropriate because it illustrates how managerial difficulties, human 'error' and technical security failures combined to create systems weaknesses. The account used in this paper draws heavily on the report to AIB by the Promontory Financial Group and by Wachtell, Lipton, Rosen and Katz (Promontory, 2002). Other sources have also been used and these are acknowledged at the point at which their material is introduced.

In 1983, the Allied Irish Bank (AIB) acquired a stake in Allfirst, then known as the First Maryland Bancorp. This stake grew until by 1989, AIB had taken acquired First Maryland through a merger. AIB planned to diversify its operations in North America. They believed that this could best be achieved by allowing Allfirst a large amount of local autonomy. Allfirst continued have its own management team and board of directors. However, stronger control was retained over Treasury operations via the appointment of a senior AIB executive to oversee these operations. Prior to his appointment in 1989, there had only been a minimal history of currency trading at Allfirst with limited risks and a limited budget. In 1990, however, a trader was recruited to run proprietary trading. These operations continued relatively successfully until the first incumbent of this post had to be replaced in 1993. John Rusnak was recruited from a rival bank in New York, where he had traded currency options since 1989. One aspect of his recruitment was the desire by Allfirst to exploit a form of arbitrage that Rusnak specialized in. This took advantage of the differences in price between currency options and currency forwards. In simple terms, an option is an agreement that gives the buyer the right but not the obligation to buy or sell a currency at a specified price on or before a specific future date. If it is exercised, the seller must deliver the currency at the specified price. A forward is a contract to provide foreign exchange with a maturity of over 2 business days from the transaction date.

Rusnak's activities can be seen in terms of the primary violations described in Figure 1. He created bogus options to hide losses that he had sustained in currency trading. These catalytic events exploited underlying vulnerabilities, similar to those sketched in Figure 2. For example, the immediate report into the fraud identified 'numerous deficiencies' in the control structures at Allfirst. In line with Mackay's assertions about causal complexes, the report went on to argue that 'no single deficiency can be said to have caused the entire loss' (Promontory, 2002). The underlying vulnerabilities included the failure of the back-office to confirm Rusnak's bogus options with the counterparties involved in the transaction. Such checks might have revealed that these counterparties had no knowledge of the fictitious transactions that Rusnak said they were involved in. .

Many of the secondary problems at Allfirst relate to their organizational structure. Allfirst's treasury operations were divided into three areas. Rusnak's currency trading was part of the front office. The middle office was responsible for liability and risk management. The back-office was responsible for confirming, settling and accounting for foreign exchange and interest rate derivatives trades, including those initiated by Rusnak. Allfirst espoused the policy of having the back-office confirm all trades, following industry practice. The initial reports speculate that Rusnak may have put pressure on his colleagues not to confirm all of his options trades. Figure 4 sketches the relationship between the different reporting structures in the Allfirst treasury. Rusnak formed part of a relatively small and specialized group in the Foreign Exchange area. This diagram also illustrates some of the potential vulnerabilities in the reporting mechanisms within the bank. The Allfirst Tresurer was responsible both for ensuring profitable trading and for ensuring effective controls on that trading. Subsequent investigations also revealed concerns about the Treasury Funds Manager's position. Not only did they direct many of the Treasury operations but they also controlled many of the reporting procedures that were used to monitor operational risks. The Vice President for Risk Control, therefore, devised a plan so that asset and liability management reports as well as risk control summaries would be directed to senior management through his office. Unfortunately, this plan does not seem to have been implemented before the fraud was detected.

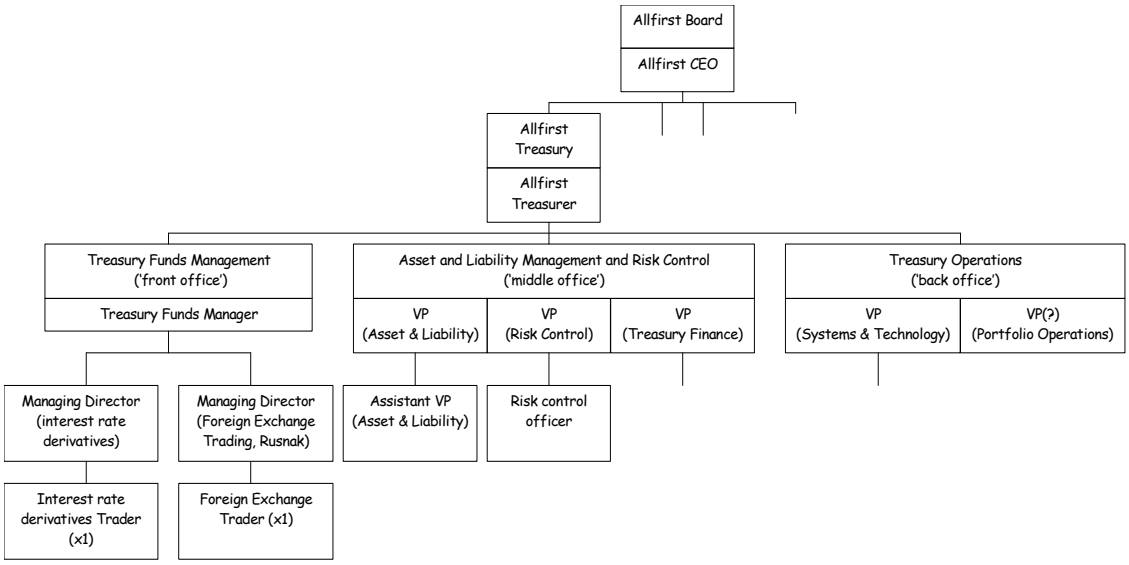


Figure 4: High-level Overview of the Allfirst Management Structure

The previous paragraphs have summarized the primary violations and secondary vulnerabilities that contributed to the Allfirst fraud. The failure to investigate potential security issues once they had been identified also illustrates tertiary failures of the type described in the opening sections of this paper. The main aim behind this overview has been to provide a concrete example of the complexity of causal arguments in security incidents. The following sections use this initial analysis to illustrate how root cause analysis techniques can be extended from accident investigations to examine a wider class of security failures.

INTRODUCTION TO ROOT CAUSE ANALYSIS TECHNIQUES

Root cause analysis techniques provide tools for identifying the elements of a causal field from a mass of other contextual factors. In Mackay's terms they can also be used to determine the composition of various causal complexes within such a field of relevant factors. Recall that each causal field is one of several possible combinations of factors that might lead to an adverse outcome. Each individual factor within a field is necessary but, typically, not sufficient for an incident to occur. As previous sections have argued, without appropriate tools it is likely that analysts will miss important factors within one or more of these causal fields. It is also likely that individual differences will lead to inconsistency between the findings of multiple independent investigators. In other words, there are likely to be significant differences over whether or not a particular factor is a necessary cause of an adverse event. This can be illustrated by the subsequent debate and litigation as to whether the prime brokerage accounts played a significant role in the causes of Allfirst's eventual loss. Root cause analysis techniques provide tools and techniques that can be used to encourage agreement over those factors, violations and vulnerabilities, that contribute to a security failure.

Barrier Analysis

The previous summary of the Allfirst fraud provides a false impression of the problems that face investigators in the aftermath of a security violation. The outcome is often, but not always, fully understood. Far less is known about the vulnerabilities that created the context for particular violations. In consequence, most investigations begin with a prolonged period of elicitation where evidence is gradually gathered about the course of an incident. Barrier analysis can be used to support these parallel activities. It also provides documentary evidence to help demonstrate that investigators have considered a broad range of causal fields.

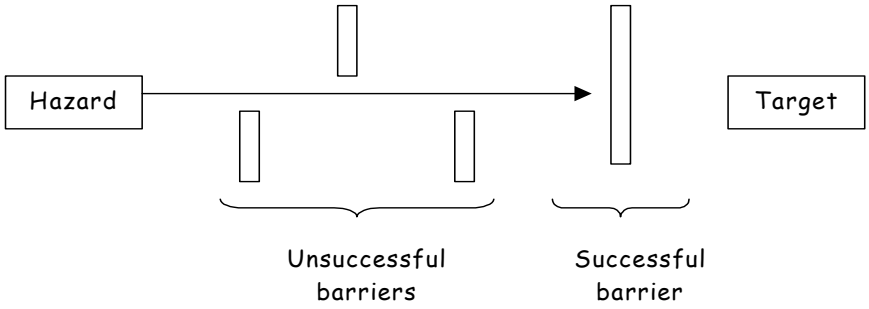


Figure 5: Targets, Hazards and Barriers

Barrier analysis is based on the idea that most security-critical systems rely on counter-measures or barriers that are intended to prevent a security hazard from gaining access to or adversely affecting a target. Figure 5 provides an overview of the central ideas in Barrier Analysis. As can be seen, a security hazard must pass through a series of potential barriers before they can reach the ultimate target. The weaknesses in these various barriers can be seen as the vulnerabilities mentioned in previous sections. The events that undermine these barriers have been called violations. In Figure 5, the final barrier denies access to, or prevents the security hazard from affecting, the target. This typifies the way in which a final layer of defenses can make the difference between an unsuccessful attack and a security breach. In such circumstances, incident investigations provide important insights both about those barriers that failed and those that acted to protect the target from a security hazard.

What?	Rationale
Hazard	Currency trading losses concealed by fraudulent use of the Bank's assets.
Targets	Allfirst's risk exposure and ultimately the Bank's assets.

Table 1: Hazard and Target Identification

Table 1 illustrates the initial stages of a barrier analysis. Investigators must first identify the hazard and targets involved in a security incident. During these initial stages, the analysis is conducted at a relatively high level of abstraction. The investigation progresses by examining the barriers that might prevent a hazard from affecting the targets. Analysts must account for the reasons why each barrier actually did or might have failed to protect the target. Table 2 illustrates the output from this more detailed stage of analysis. As can be seen, the barriers are those defenses that were intended to prevent undisclosed currency trading losses from distorting the bank's risk exposure and reducing the Bank's assets. As can be seen, the Value at Risk calculation (VaR) was one of the barriers that were intended to protect Allfirst's risk exposure. The policy of ensuring that the back-office confirmed all transactions should also have kept bogus trades off the balance sheet. Similarly, the foreign exchange rates used by the middle and back-office should have come from an independent source to reduce the risk of manipulations masking the true risk of any options. Analysis progresses by identifying the reasons why each of these barriers was breached. In other words, this approach helps to identify the vulnerabilities that might have been exploited in the course of any security incidents. The development of such a table also helps to focus the gathering of evidence. Investigators must obtain proof that supports the assertions made in Table 2. For example, the Value at Risk calculation provided Allfirst with a statistical measure for the 'worst case' losses that might be expected from a given portfolio. Monte Carlo techniques were used on a range of hypothetical spot and volatility rates for different currencies. The VaR was derived from the tenth worst outcome obtained from the stochastic model. However, as can be seen in Table 2, Rusnak distorted the VaR calculations by creating bogus options that hedged, or mitigated, the risks associated with his real trades. Similarly, the VaR calculation was distorted by the manner in which Rusnak persuaded an employee in the Risk Control Group to use data from his PC to calculate the extent of 'holdover transactions'. Rusnak often held very large 'open' positions. In general terms, an open position is a deal that is not yet reversed or settled. The investor is subject to exchange rate movements. Hence, the larger the open position then the greater the risk denoted by the VaR. Any transactions that could not be fully processed at the end of the

day were 'heldover' until the following morning. These were not entered into the bank's trading software and so were not included in the VaR calculations. It would have been relatively easy to check whether trades were being hidden. This could be done by adding them to the 'holdovers' and then removing them the next day. For example, internal audit might have compared the list of holdovers against the next day's trading activity to ensure that the trades had, in fact, been entered into the trading software. The abuse of the holdover system was so extreme that several transactions were left in the system for three consecutive days without being entered into the trading software. The manner in which Rusnak was able to enter false transactions into the DEVON system represents a further way in which the VaR barrier was undermined. This software was used to record all transactions on the prime brokerage accounts. In particular, he entered false forward transactions and then reversed them prior to the next settlement date. This enabled him to create transactions that would alter the VaR, for instance by hedging other transactions. These bogus trades could then be removed from the system before any settlement took place.

Barrier	Reason for failure?
Value at Risk Calculations	Rusnak created bogus options to hedge his real positions and hence those options offset or reduced the Value at Risk (VaR) from his activities.
	Rusnak provided false figures for 'holdover' transactions. An employee in the Allfirst risk control group used a spreadsheet to calculate the VaR that relied on files obtained from Rusnak's PC.
	Rusnak was able to enter false foreign currency exchange forward transactions in the DEVON system and reverse them before the settlement date, manipulating the transactions that were used to calculate the VaR.
Back-office confirmation of bogus options with counterparties.	Back-office staff did not confirm all of Rusnak's transactions. Promontory group report refers to bullying and cajoling.
	Difficulty of confirming Asian trades in middle of US night persuades treasury senior staff to decide not to confirm offsetting pairs of trades with these markets.
Foreign exchange rate confirmation from independent sources.	Rusnak opposes direct drop of Reuters prices into middle and back-office systems.
	Risk assessment analyst fails to communicate concerns to senior management.
	Decision not to invest in dedicated Reuters terminals for middle and back-office staff.
	Delays in implementing back-office 'computer disk' to store prices direct from Reuters once senior management alerted to downgrade in second quarter 2001 risk assessment report.

Table 2: More Detailed Barrier Analysis

Table 2 identifies a number of other barriers that were intended to protect Allfirst's risk exposure from bogus transactions. In particular, it is a common assumption that back-office staff will confirm that all trades have actually been made or placed with counterparties. Rusnak was able to create bogus hedging positions because this did not happen at Allfirst. The subsequent investigations revealed that Rusnak had pressurized responsible staff in the Allfirst treasury so that they did not routinely check all of his transactions. In addition to such persuasion, a number of other reasons explain this apparent violation of company policy. Firstly, Rusnak typically created pairs of offsetting transactions so it looked as though the risk in one trade was matched by hedging positions in another. The lack of any cash payments and the problems associated with contacting traders in the Asian markets may have combined to reassure the staff member involved that confirmation was not always necessary.

A final barrier mentioned in Table 2 is the practice of obtaining currency exchange rate information from an independent source. In the past this had been done by printing the exchange rates directly from the Reuters feed to the front office and then cross checking these values against the rates used by the middle and back-offices. The treasury risk control analyst argues that this was error prone and cumbersome and that a spreadsheet might be developed to speed access across these departments. The intention was that the software would directly interface with the Reuters link. However, this potential barrier to exchange manipulation was undermined by Rusnak's insistence that he have access to the spreadsheet so that he could use the exchange rates to monitor his VaR. The treasury risk control analyst, therefore, allowed an architecture to develop whereby the Reuters feed to the front office was downloaded onto Rusnak's PC. The middle and back-office system then drew data from these files on Rusnak's machine. The risk

assessment analyst subsequently met with the treasury risk control analyst and queried whether this was appropriate. Table 2 also captures the observation that 'the risk assessment analyst failed to communicate concerns to senior management'. Senior management did eventually become aware of this procedure when the risk assessment group downgraded the first quarter 2001 risk assessment report from Good to Weak. However, the Barrier Analysis also records that there was a significant delay before the back-office was equipped with their feed and disk to store the Reuters currency information.

As can be seen, the barrier analysis represented in Figure 5 encourages analysts to consider both the underlying vulnerabilities and violations that combine to compromise the security of many complex systems. For instance, Rusnak's manipulation of the 'holdover' transactions was only possible because there was an underlying vulnerability created by the failure to check that such trades had actually been entered during the next working day. Similarly, Rusnak's manipulation of the Reuter's feed was only possible because of the decision not to provide middle and back-office staff with their own dedicated links.

Change Analysis

Change analysis provides a similar form of support to that offered by barrier analysis. Rather than focusing on those defenses that either protected or failed to protect a potential target, change analysis looks at the differences that occur between the actual events leading to a security incident and 'ideal' operating procedures. For example, the actual mechanisms used to obtain pricing information might be compared with those described in a company's risk control documents. Table 3 provides an example of change analysis. The first column describes the ideal condition. In some applications of this technique, the first column is instead used to represent the practices and procedures that held immediately prior to a security incident. This is an important distinction because the causes of an adverse event may have stemmed from inappropriate practices that continued for many months. In such circumstances, the change analysis would focus less on the conditions immediately before the incident and more on the reasons why practice changed from the ideal some time before the mishap.

Table 3 shows that Rusnak's supervisors should have examined his positions and trades in greater depth given the overall size of his positions. This 'normative' statement can be justified by referring to a range of Allfirst and AIB documentation on internal audit and risk control (Promontory, 2002). The middle column indicates several of the ways in which Allfirst practice differed from this norm. No one noticed that many of Rusnak's options expired unexercised on the day that they were created. This enabled him to leave bogus balancing transactions on the book. The longer-term bogus transactions avoided suspicion because they appeared to be hedged by the short-term options that expired unexercised. Normal security precautions such as telephone tapping and logging were not used. This deprived risk control managers of important sources of information that might have alerted them to the lack of communication with the counterparties on many of the bogus options. Finally the lack of scrutiny on Rusnak's positions is revealed by the failure to reconcile his daily profit and loss figures with the general ledger at Allfirst. One consequence of this was that Rusnak was able to develop trades well beyond his daily limits, for example by the abuse of the holdover system mentioned in previous sections.

Prior/Ideal Condition	Present Condition	Effect of Change
Rusnak's supervisors should have examined in depth his positions and trades given the overall size of those positions.	No one in Allfirst noticed the options that expired unexercised on the day they were created.	Rusnak was able to create bogus options because he created two balancing transactions, the first would expire the next day unexercised but the second would remain on the books offsetting apparent losses.
	Normal precautions like telephone tapping and data logging were not used.	This might have revealed the lack of calls or other communication with counterparties on bogus options.

	There was no reconciliation of Rusnak's daily profit and loss figures with the general ledger.	The generation of bogus options created large daily volumes in excess of the limits normally placed on Rusnak's transactions. The lack of reconciliation prevented the identification of several of the bogus transactions such as the "holdovers" that were never entered on the general system.
A process of internal audit should ensure that suggestions made by audit, risk assessment and supervisory examinations are fully followed through.	Several recommendations were acted on but others were not and there seems to have been no systematic process for recording that urgent or important actions received adequate review.	Several reports document the dangers of not ensuring independent sources for currency information. There was some delay in following up these reports even when the problem was recognized. Rusnak used these vulnerabilities to hide his losses, for instance through the VaR calculations

Table 3: Change Analysis

Table 3 also illustrates the argument that normal auditing practice should ensure that suggestions made by audit, risk assessment and supervisory examinations are followed through until they are either implemented or reasons for their rejection are adequately documented. In contrast, several recommendations were ignored or only implemented in a piecemeal fashion during the Allfirst fraud. The lack of systematic monitoring for auditing recommendations created opportunities for Rusnak. The resulting vulnerabilities included a considerable delay in establishing independent sources for currency pricing information. This enabled Rusnak to manipulate the VaR calculations for his trading activities.

An important benefit of change analysis is that the 'ideal' conditions in these tables can be used to identify recommendations. This is not straightforward. For instance, stating that staff and management should follow the company's risk control procedures does not provide any guarantee of compliance. The prior/ideal condition column in the change analysis tables can, however, provide a starting point for the identification of more detailed recommendations. In Table 3, investigators might argue that a monitoring system should be introduced to trace the implementation of audit, risk assessment and supervisory examinations. It should then be possible for senior management to use the system to ensure the implementation of necessary interventions recommended by these internal audits. Had such a system been adequately implemented then Allfirst might have avoided or minimized the delays associated with the development of an independent currency pricing system from the middle and back-offices.

A number of limitations restrict the utility of change analysis. For instance, they often introduce a form of hindsight bias. Norms were not followed because violations were able to exploit existing vulnerabilities. It is, therefore, tempting to argue that existing rules and regulations should be applied more diligently in the future. This is a dangerous argument. It assumes that existing procedures and practices were sufficient to ensure the security of a system. Further limitations affect both Barrier Analysis and Change Analysis. These techniques can be used to structure the initial analysis of a security incident. They guide investigators by providing a framework of important concepts as they gather information about what should have happened and what actually did occur during particular violations. They do not, however, provide more detailed support for the modeling that is often necessary in order to understand the complex manner in which different events and causal factors combine over the course of a security incident. Event based modeling techniques can be used to avoid this limitation during the reconstruction of complex failures.

VIOLATION AND VULNERABILITY ANALYSIS (V^2 ANALYSIS)

Many different event-based techniques have been developed to support the root cause analysis of safety-related incidents. These include Events and Causal Factors charting (ECF), Multilinear Events Sequencing (MES) and Sequential Timed Event Plotting (STEP). Brevity prevents a detailed analysis of each of these approaches; the interested reader is directed to Johnson (2003). The key point is that several of these techniques have also been used to analyse the underlying vulnerabilities and specific violations that lead to security related incidents (US Department of Energy, 1992). Most previous applications have been within the specific context of nuclear energy and weapons development. A further limitation to the more general application of these techniques is that they provide little specific support for the analysis of security incidents. Hence, the basic components in these event-based techniques are unchanged from their use in

safety-related applications even though the details surrounding these 'dependability' failures can be very different.

In contrast, Figure 6 provides an example of Violation and Vulnerability (V^2) analysis. This extends an event based modelling technique to deliberately support the identification of root causes for a wide range of security related incidents. The underlying approach is similar to the existing ECF, MES and STEP techniques, mentioned above. This V^2 diagram is constructed around a number of events that are denoted by rectangles. For example, 'AIB insert senior manager as Allfirst treasurer' and 'Treasurer is appointed to key AIB group marketing strategy committee' are both shown as events in Figure 6. These are made more likely by a number of contributory factors that are shown by ellipses. For instance, the decision to insert one of the AIB executives as the Allfirst Treasurer led to a situation in which some viewed the treasurer as a form of 'home office spy'. This contributed to the exclusion of the formed AIB executive from some senior management decisions at Allfirst.

Figure 6 focuses more on the contextual factors than on specific events during the Allfirst fraud. It also maps out a range of conditions that formed the background to the more detailed events that are mentioned in previous sections. This is deliberate because an important objective behind the use of this modeling technique is to trace the roots of a security violation back into the underlying vulnerabilities within the operations of a company, such as Allfirst. Vulnerabilities can be thought of as a particular type of contributory factor. As mentioned in Figures 1 and 2, they create the opportunity for the violations that occur during security incidents. In Figure 6, vulnerabilities relate to the dual reporting structure between AIB and Allfirst. They weakened the supervision of the Treasurer's activities in the lead-up to the fraud. This vulnerability is denoted by the double ellipse at the bottom right of figure 6. Subsequent V^2 diagrams can be used to map out the precise manner in which this particular contributory factor acted as a precondition for Rusnak's violations.

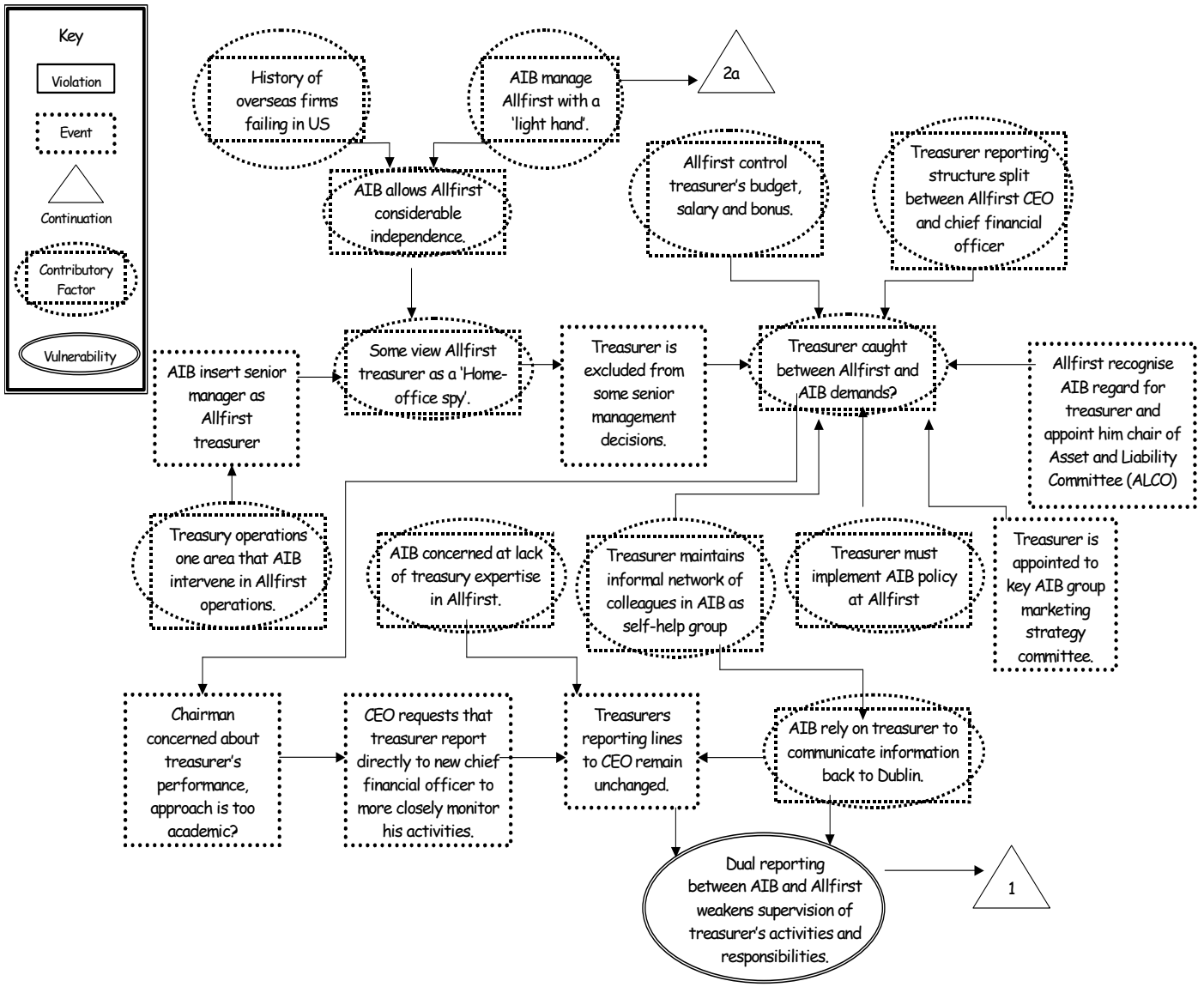


Figure 6: A V^2 Diagram of the Background to the Allfirst Fraud

Figure 6 illustrates the way in which V^2 diagrams can be used to look beyond the particular violations that lead to a fraud. This is important if investigations are to accurately identify the underlying managerial and organizational factors that might lead to future security problems. For instance, one response to the events at Allfirst would simply have been to focus legal retribution on the trader. This would, however, have ignored underlying problems in the relationship between AIB and Allfirst, including the supervision of key Treasury staff. This point is made more forcefully in the recommendations that emerged in the immediate aftermath of the fraud; 'In light of the foregoing considerations, AIB should consider terminating all proprietary trading activities at Allfirst, and all customer trading activities at Allfirst should be relocated to the AIB branch in New York. While the salespeople may continue to be located in Baltimore, any price-making and trade execution should be done in New York, under the direct supervision of AIB treasury' (Promontory, 2002).

Figure 7 continues the Violations and Vulnerability analysis by documenting the events leading to the hiring of Rusnak by Allfirst. Prior to 1989, Allfirst had only engaged in limited currency trading. This contributed to the decision to recruit a specialist to run their proprietary trading business. During this period, trading was focused on directional trading, in other words profits were dependent on forecasting the future price of a currency as it moved up or down on the markets. The senior trader left Allfirst and a further event in Figure 7 is used to show that the 'Treasury funds manager heads the search for a new trader'. This leads to an offer being made to Rusnak. The decision to make this offer was supported by recommendations from his previous employers at Chemical Bank. His appointment was also supported by the Allfirst Senior Management's interest in Rusnak's non-directional trading. This will be described in more detail in subsequent V^2 diagrams. Figure 7 also illustrates how these various events, together with a number of additional contributory factors lead to a further security vulnerability. Allfirst's efficiency committee suggested that the treasurer scale-back proprietary currency trading. However, the senior management interest in Rusnak's non-directional approach helped to focus the cutbacks in more conventional forms of currency trading. The senior management interest also created a situation in which the Treasury funds manager was highly protective of Rusnak and his activities. These various factors combined to weaken the monitoring and reporting procedures that were established to control the risks associated with his activities. When Rusnak's immediate trading manager resigned, his post was not filled. Lack of funds prevented a renewed appointment and so Rusnak now reported directly to the treasury funds manager who, as we have already seen, was protective of his non-directional trading strategies.

Analysts can use V^2 diagrams to map out the mass of contextual details that emerge during an investigation. Change and Barrier analysis can be used to identify these contributory factors and events. A number of other approaches, such as Conclusion, Analysis and Evidence diagrams and Why-Because Analysis, have been developed within the field of accident analysis to exploit more narrow definitions of causal relationships than those illustrated in Figure 7 (Johnson, 2003). Alternatively, Multilinear Event Sequencing (MES) is one of several techniques impose additional formatting constraints on diagrams that are similar to those shown in this paper (US Department of Energy, 1992). MES uses a grid in which the events relating to particular actors or agents had to be shown along the same row. Columns were then use to denote the flow of events over time. Each event had to be shown to the right of the events that occurred before it. In contrast, V^2 diagrams take a more relaxed approach. It can be difficult to establish the exact timing for many events. This problem can be even worse for contributory factors. For instance, when should an investigator show that 'Senior management were intrigued by Rusnak's non-directional trading approach'? This sentiment seems to have emerged over a prolonged period of time and cannot easily be associated with particular meetings or events, especially in the aftermath of a security incident. Similarly, other events affect many different actors in an adverse event. In Figure 6, several different managers supported the appointment of the Treasurer on the AIB and Allfirst committees. These events would have to be widely distributed across many different columns in a MES diagram adding to the complexity of constructing and maintaining these representations. It is for this reason that V^2 diagrams relax some of the constraints that guide Multilinear Event Sequencing. Arrows represent relationships or constraints. They do not represent necessary causal relationships. For example, the protective attitude of the Treasury Funds Manager did not 'cause' the flaws that affected the reporting and monitoring of Rusnaks work. The fraud may even have occurred if the Treasury Funds Manager had not been so protective. However, the manner in which he shielded the trader from subsequent enquiries did have a profound impact on the underlying vulnerability illustrated in Figure 7.

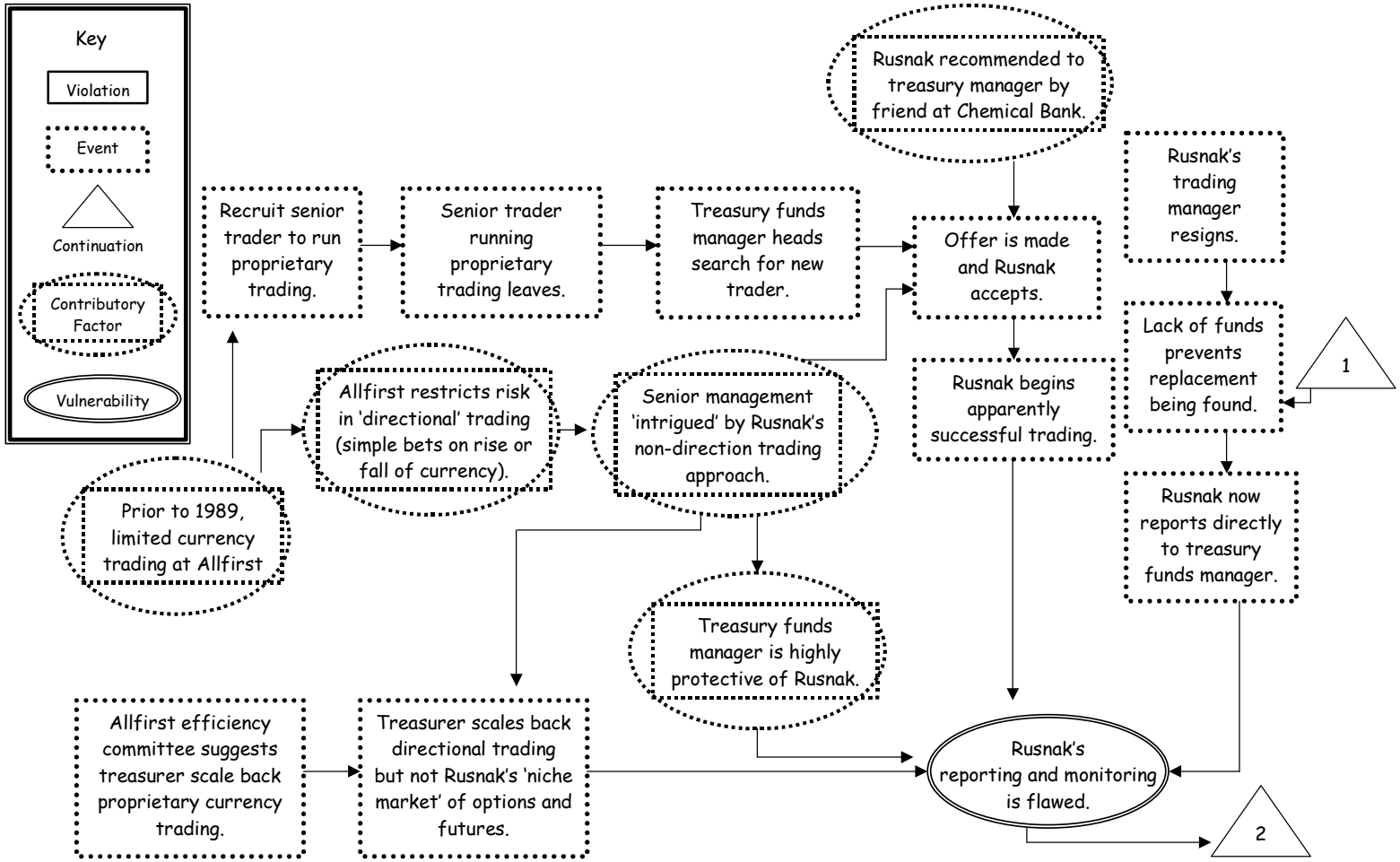


Figure 7: A V^2 Diagram of the Events Leading to Rusnak's Appointment and Flaws in his Reporting Structure

Figure 8 extends the V^2 analysis towards the events surrounding Rusnak's fraudulent activities. As can be seen, he initially created the impression that he specialized in a form of arbitrage by taking a profit from differences in the exchange rates between different markets. In particular, he claimed to make profits by holding a large number of options that were hedged by balancing positions in the cash market. These observations are denoted in Figure 8 by the contributory factors at the top-right of the diagram. The contributory factors at the top-left show that most of his trades were simpler than many at Allfirst had supposed. They involved linear trades based simply on predicted fluctuations in currency rates. This led him to buy significant quantities of Yen for future delivery. The subsequent decline in value of this currency prior to delivery left Rusnak with a loss. Combined with the image that he had fashioned for his trading activities, the loss may have created a situation in which he felt under pressure to hide the outcomes from his options on the Yen. This analysis of the top components in Figure 8 raises a number of important issues about the construction of V^2 diagrams. It can be argued that Rusnak's creation of a false impression about the nature of his trades should be 'promoted' from a contributory factor to either a violation, and therefore be linked to specific events, or vulnerability. The tension between his claimed trading techniques and his actual methods helps to explain many of his subsequent actions. It can equally well be argued that such tensions are widespread within many financial organizations. Several studies have pointed to the psychological characteristics and personality attributes of successful traders (Tvede, 1999). It has been argued, for instance in Oberlecher's (2004) study of the psychology of foreign exchange markets, that the same attributes that create these tensions between action and appearance may also be important ingredients in the makeup of successful traders. The meta-level point here is that V^2 analysis forces investigators to consider whether or not each contributory factor could be considered a potential vulnerability and also whether each event in the context of a security incident might also be labeled a violation. There is no automatic or algorithmic process to support this analysis.

Figure 8 also illustrates the mechanisms that Rusnak used to hide his losses from directional trading on the Yen. These have been briefly outlined in previous sections. Initially, he began by creating a bogus 'deep in the money' option. Recall that such an option has a price that is significantly below the current spot-price and hence it is high risk for the vendor. Such options attract high premiums, especially if they can be exercised in the short term when the spot price is unlikely to fall below the level of the quoted option. Allfirst, therefore, had a significant potential liability. At the same time, he created a second balancing bogus option with the same counterparty. This is represented in Figure 8 by the violation labeled 'Rusnak creates balancing option as if Allfirst have paid a large premium to buy currency weeks later involving the same counterparty'. This made it look like Allfirst's original liability was offset by the asset value of the second option. Allfirst should have paid a correspondingly large premium to obtain this second option even though no cash would actually have changed hands because the two premiums balanced each other and were drawn against the same parties. The crucial difference between these options was that the first one, representing Allfirst's liability, was set up to expire within 24 hours. The second, representing Allfirst's fictitious asset, expired several weeks later. Rusnak knew that neither option would ever be exercised because they were bogus deals. However, for the period between the expiry on the first option and the end of the second, he was able to create the appearance of a genuine asset on the Allfirst books. This could be used to offset his own genuine losses.

These deals made no sense for a number of reasons. Firstly, the risk exposure on each of the options was quite different given that one expired in 24 hours while the second typically lasted for several weeks. In such circumstances, the options should have attracted very different premiums and so were unlikely to balance each other out. Secondly, the 'deep in the money' options involved in the first bogus trade should have been exercised by the counterparty. A series of similar options failing to be acted upon should have alerted company management to potential fraud. However, as Figure 8 also shows, Allfirst managers did not have access to a list of those options that had expired without being exercised within 24 hours of them being placed. This is denoted by the vulnerability on the left hand side of the V^2 diagram. Prior to September 1998, Rusnak covered his tracks by creating bogus confirmations from the supposed counterparties to these transactions. The confirmations were intended to provide evidence that both parties had agreed upon these trade options. After that time, Rusnak managed to persuade the back-office staff not to pursue these confirmations for his trading activities. As can be seen from the V^2 diagram, their failure to confirm the transactions is partly explained by the difficulty of establishing contact with many of Rusnak's brokers who worked in the Asian offices of the counterparties. The trader's office hours often created considerable communications difficulties for Allfirst's back-office staff. Figure 8 also uses a triangle continuation symbol, labeled with a '2', to carry the analysis from the events surrounding Rusnak's appointment to the start of his fraud. As can be seen, flaws in the reporting and monitoring procedures for Rusnak's activities made it more likely that he would be able to persuade back-office staff not to confirm the matching pairs of bogus trades. These flaws stemmed in part from senior management's desire to support his 'novel' forms of arbitrage.

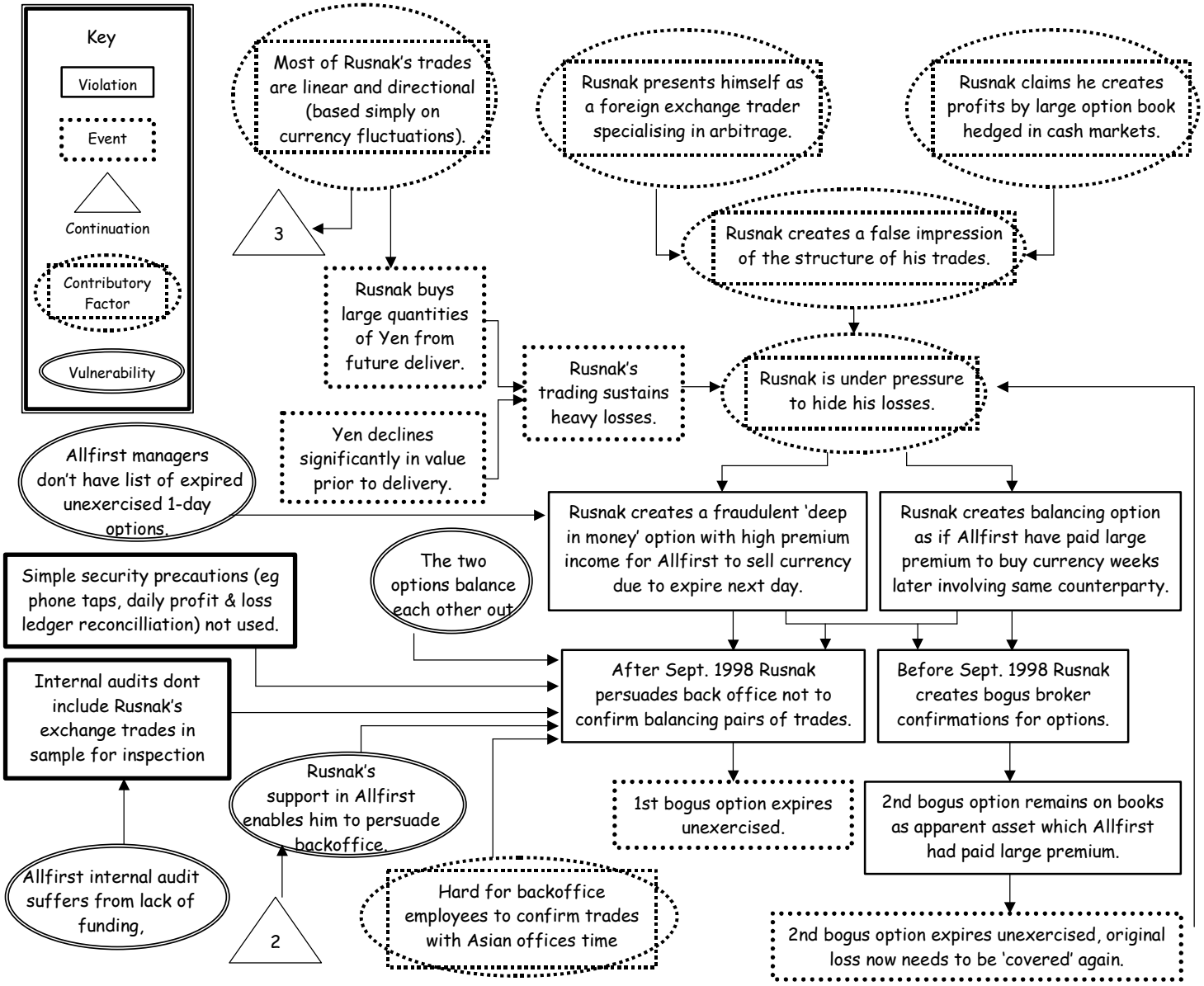


Figure 8: A V^2 Diagram of Rusnak's Initial Balanced-Options Fraud

Figure 8 also captures the cyclical nature of Rusnak's fraud. Eventually the second option in each bogus pair would expire unexercised. At this point, the large and fictitious asset would disappear from Allfirst's books. It was, therefore, important that Rusnak continue to generate these option pairs if his fraud was not to be discovered. This is indicated by the arrow from the bottom right of Figure 8 between '2nd option expires unexercised, original loss now needs to be covered again' back to the contributory factor 'Rusnak is under pressure to hide his losses'. In previous V^2 diagrams, rectangles have been used to denote specific events. In contrast, Figure 8 shows the structure of Rusnak's fraud using generic events that represent a class of similar violations. It would, of course, be possible to construct a more specific model that represents each of the individual trades that made up this pattern within the security incident. However, the level of detail illustrated in the previous diagram is appropriate for most stages of an investigation. At this stage of the analysis, there is little to be gained from individually identifying the unwitting counterparties to Rusnak's options trades.

Figure 9 continues the V^2 analysis of the Allfirst fraud. The ability to represent change over time is important because many security incidents develop over months or years. The individuals and groups involved often alter their behavior in response to external events and the audit mechanisms that are used to detect any continuing vulnerabilities. This diagram shows how Rusnak exploited further opportunities to expand both his trading activities and the range of bogus trades that were required to conceal his mounting losses. The top right event in Figure 9 denotes that Rusnak was offered net settlement agreements with a number of financial institutions (Promontory, 2002). These eventually developed into 'prime brokerage accounts'. Such facilities enabled the broker to settle spot foreign exchange transactions with the counterparties. Each of these individual trades was then rolled together into as larger forward transaction between the broker and Allfirst that could be settled on a fixed date every month. As can be seen, these agreements simplified multiple transactions between Allfirst and the counterparties into a smaller number of larger transactions with the brokers. This simplification had two effects. Firstly it reduced the number of operations for the Allfirst back-office. Secondly, it made it difficult for the back-office and others within Allfirst from monitoring the individual trades that were being roller together within Rusnak's prime brokerage accounts. This potential vulnerability is represented half way down Figure 9 on the right hand side.

The problems of monitoring transactions through the prime brokerage accounts together with the ability to roll together individual transactions for periodic settlement together combined to create a situation in which Rusnak could exceed the limits on his trading that were routinely insisted upon by Allfirst. His ability to increase the scope and scale of his trading is shown in Figure 9 to have increased the amounts of his loses in both forward and spot transactions. In order to cover his losses, another cycle emerged in which he generated more bogus transactions using the balancing options approach, described in previous sections. Rusnak was also able to exploit vulnerabilities in the DEVON software. This was used to track trades across the prime brokerage accounts. He was able to enter bogus transactions into the system and then reverse them before the monthly settlement period. As can be seen, however, Figure 9 does not provide sufficient details about the nature of the underlying problems with the DEVON application. The vulnerability symbol is annotated with the comment; 'DEVON system vulnerabilities (further analysis?)'. The V^2 notation could be revised to explicitly represent this need for additional analysis. More symbols could be used to show those events and contextual factors, violations and vulnerabilities that have only been partially analyzed. This has not been done, however, in order to minimize the amount of investment that must be made in training to both read and eventually develop these diagrams.

The right-hand, lower portion of Figure 9 illustrates a series of events that threatened Rusnak's activities. It began when the Allfirst treasurer decided to introduce a charge on those activities that used the bank's balance sheet. Such a change would provide greater accountability, for example by exposing whether the profits generated by an activity actually justified the work created for those who must maintain the balance sheet. Questions began to be asked about whether the apparent profits from Rusnak's activities could justify his use of the balance sheet. The total volume of currency traded had risen rapidly over the year to January 2001 but net trading income remained almost the same. A significant proportion of this rise can be attributed to Rusnak's various trading activities. He was, therefore, told to reduce his use of the balance sheet. This not only curtailed his legitimate trading activities but also placed tight constraints on many of the bogus trades, even if many of those trades only made a fleeting appearance on the Allfrist books before being reversed. He had to identify an alternate source of funds to offset his previous losses and those that continued to accrue from his legitimate trading activities.

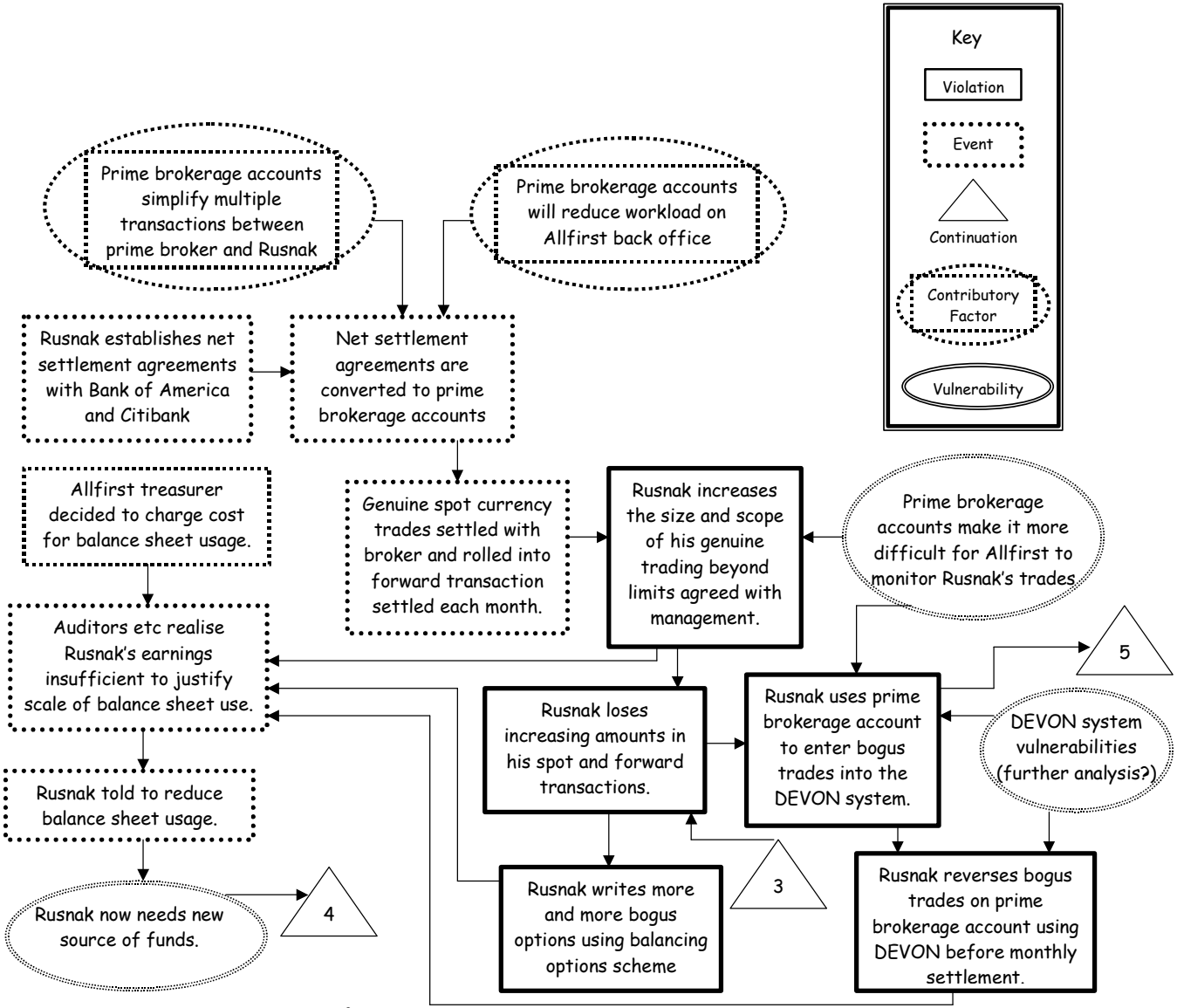


Figure 9: A V^2 Diagram of Rusnak's Manipulation of Prime Brokerage Accounts

Figure 10 traces the Allfirst fraud from the point at which senior management began to question Rusnak's use of the bank's balance sheet. This is denoted by the continuation symbol, labeled 4, connecting this image with the V^2 diagram in Figure 9. Rusnak's need to find an alternate source of funds led him to sell long-term options that were deep in the money. As mentioned previously, these options quoted a strike price that was far above the currency's current spot price. Hence, the options represented a relatively high-risk for Allfirst and attracted a corresponding premium. However, Figure 10 also uses a contributory factor to denote that these 'deep in the money options can be viewed as a form of loan' and that 'Rusnak would need to get these liabilities off the books'. Allfirst would have to redeem them when the options were redeemed. Figure 10 denotes a further violation as Rusnak created bogus transactions to indicate that the original options had been repurchased. These activities again involved Rusnak's use of the balance sheet and so the Allfirst treasurer placed a limit of $150 million on his trades.

Previous V^2 diagrams have shown how Rusnak was able to manipulate the DEVON system to conceal some of his transactions via the prime brokerage accounts. Figure 10 shows some of the consequences of these manipulations through the continuation symbol, labeled 5, that links back to the previous diagram. The misuse of the DEVON system, combined with the 'bogus' repurchasing of 'deep in the money' options distorted the Value at Risk (VaR) calculations that were introduced in previous sections. Figure 10 also illustrates further ways in which this risk assessment tool was undermined. Rusnak used 'holdover transactions' to disguise some of his trades. These transactions usually occurred after it was possible for them to be included in the day's accounts. They were, therefore, held over until they could be processed during the next trading day. Internal audit and risk control were aware that Rusnak was responsible for a large number of these transactions but they did not investigate. This observation is illustrated by the vulnerability at the top right of Figure 10. Holdover transactions were not entered directly onto the bank's trading software. There were no checks to determine whether transactions were actually entered into the following day's trading. All of these vulnerabilities can be seen as causal factors in a violation of audit procedures whereby Rusnak directly supplied risk group employees with on-line data for his holdover transactions.

The VaR was the main statistical measure used by the Allfirst senior management to monitor currency trading. Figure 10 also traces other mechanisms that Rusnak used to undermine reporting within the bank. The contributory factors on the right hand side of this diagram record how currency revaluations were initially printed from a Reuters feed and then were manually checked against the back-office records. Several audits had stressed the need to ensure that this data was derived from an independent source that could not easily be manipulated by individual traders. This potential vulnerability was addressed in a treasury control analyst's proposal that Reuter's data could automatically be fed into the bank's internal systems. However, as Figure 10 illustrates, Rusnak was able to violate these security concerns by arguing that he needed constant access to changing currency rates across the bank in order for him to monitor and assess his VaR calculation. This combined with the associated expense of providing dedicated Reuter's feeds into the back-office persuaded treasury to enable Rusnak to feed data from his own software into the rest of the Allfirst systems.

The previous paragraph illustrates how V^2 analysis provides a framework for the detailed investigation of security related incidents. Rusnak's ability to persuade the treasury control analyst that it would be acceptable for him to pass on pricing information to other bank systems requires further analysis. The figure quoted for dedicated feeds and the apparent disregard of previous audits could be the focus for subsequent investigation using complementary techniques, including the Barrier and Change Analysis. The key point is, however, that these diagrams provide an overview of the complex events and contributory factors that lead to security incidents. The resulting sketches can be shown to other members of multi-skilled investigatory teams so that they are built up over time. Peer review can also help to ensure that the resulting analysis captures both the primary violations and secondary vulnerabilities that lead to adverse events.

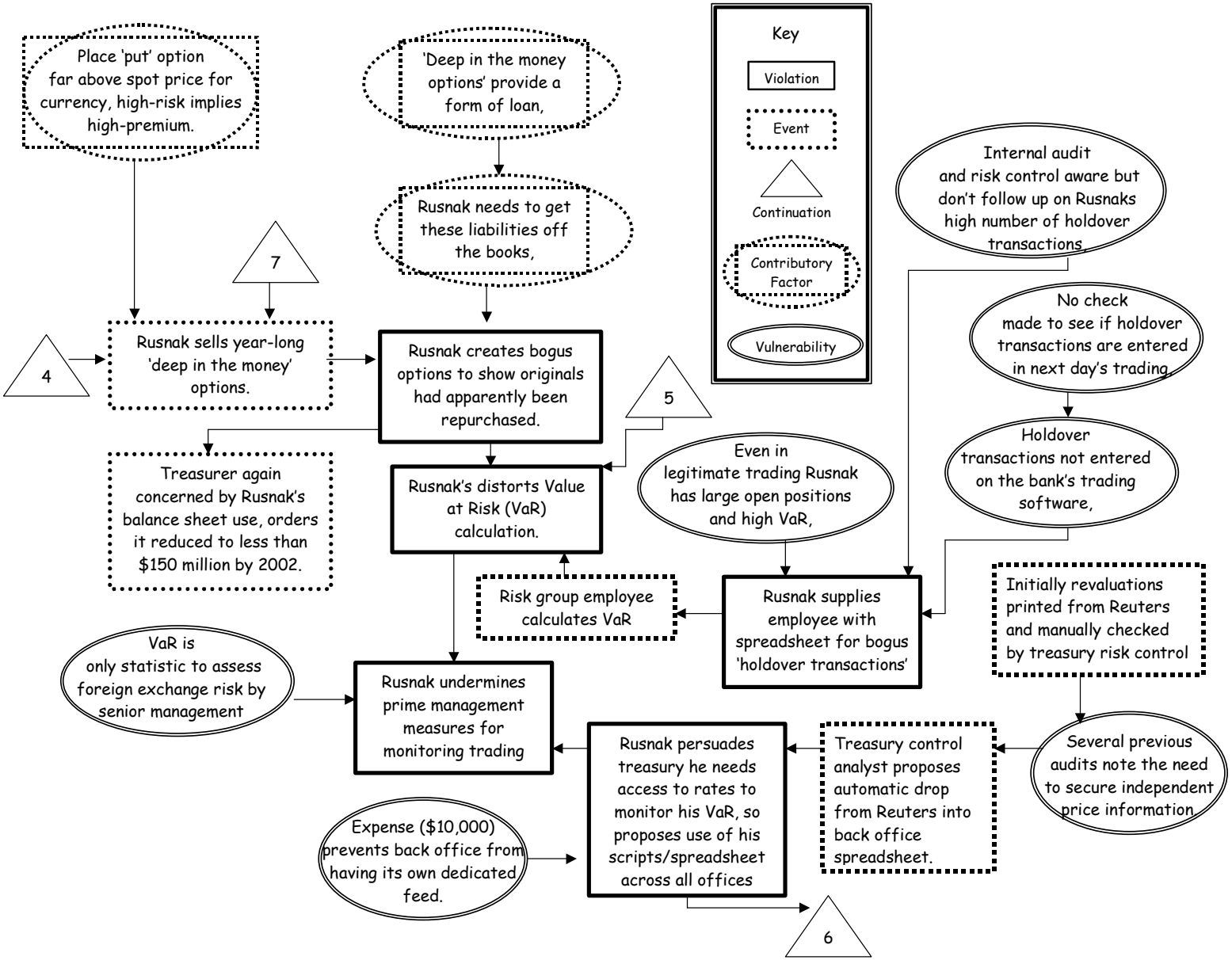


Figure 10: A V^2 Diagram of Rusnak's 'Deep in the Money' Options and the VaR Calculations

Figure 11 continues our analysis of the various opportunities that different Allfirst personnel had to detect Rusnack's activities. The continuation symbol, labeled 6, comes from Figure 10 where it was noted that Rusnack had argued to be allowed direct access to currency feed and had proposed the use of his spreadsheets and scripts by other staff. Several of his colleagues became concerned about this situation. Figure 11 carries on by denoting that a risk assessment analyst and a treasury risk control analyst met to discuss the potential vulnerabilities created by Rusnack's proposal in the previous diagram. Their meeting has three outcomes. The first is a violation 'Risk assessment analyst does not alert senior management'. Instead, the 'risk assessment analyst follows-up currency feed issues herself' and the 'treasury risk control analyst informs risk assessment that he is working on direct feed from Reuters bypassing Rusnak's software'. These last two observations are shown in Figure 11 as events rather than violations.

The identification of particular events as violations and contributory factors as vulnerabilities relies upon the subjective judgment of individual analysts. These decisions should form the focus for continued discussion within an investigation team. The outcome of this analysis is important because any further investigations are likely to concentrate on violations and vulnerabilities rather than contextual events and causal factors. For example, in Figure 11 it is important to consider the reasons why the 'risk assessment analyst does not alert senior management' to her concerns over Rusnak's control of the currency feed. In this case, the V^2 diagram shows that the 'Allfirst internal auditing department suffered from a lack of resources'. This vulnerability contributed to the violation in which serious concerns about the currency feed were not communicated to senior management because 'neither treasury specialist had experience in foreign exchange trading'. Arguably, if they had more experience then they might have been more concerned about Rusnak's access to the spreadsheets and might also have been more confident in passing those concerns up to higher levels of authority within the bank. The lack of resources had other consequences. Figure 11 shows that the treasury risk control analysts' involvement in a rerouting plan for the Reuter's feed was also the result of these limitations. Allfirst initiaily did not want to pay the additional $10,000 for a dedicated Reuter feed to the back-office. A key benefit of the V^2 analysis is that it shows how these different vulnerabilities interacted to create the context in which the fraud went undiscovered. A further benefit is that the diagrams provide a high level overview of the mass of more detailed evidence that is gathered in the aftermath of a security incident. For example, the initial investigation into the fraud concluded that:

> Allfirst internal audit appears to have suffered from inadequate staffing, lack of experience, and too little focus on foreign exchange trading as a risk area. Internal audit devotes at most two full-time auditors to auditing all of treasury. Neither of those treasury "specialists" in recent years has had a background or training in trading activities, let alone foreign exchange. The treasury audit responsibilities rest with the same team responsible for trusts (another important audit area), and the manager of that team appears to have had little trading expertise and to have done little to supervise the few treasury auditors he did have. (Indeed, this audit manager appears to have failed even to initial the work papers for the last trading audit.) Beyond audit, there are other staffing problems. The entire risk assessment department only amounts to two people who are responsible for assessing risk company-wide at Allfirst. And treasury risk control devoted only one full time employee to measuring trading risk in the foreign exchange portfolio. She was extremely inexperienced and appears to have received little support or supervision from others in treasury risk control". (Promontory, 2002, p.18)

Figure 11 could be extended to include the mass of other similar information that is available to investigators. This would, however, reduce the tractability of diagrams that are already complex. Again the decision about the level of detail to introduce into these figures must be the result of negotiation within the investigatory team. Equally, there must also be some clear mapping between the nodes in the V^2 diagram and the supporting evidence. In previous work we have done this by including unique reference numbers with each vulnerability or violation that can then be cross-references to individual documents gathered as evidence (Johnson, 2003).

The analysis of the failed barriers to Rusnak's fraud continues in the V^2 diagrams. Figure 11 also shows that one outcome of the risk assessment analyst's decision to pursue the currency feed personally was that she asked Rusnak to email her a copy of his spreadsheet that was used to pass on values to the back-office. She 'immediately discovered Yen and Euro values were corrupted' and then downgraded the control market risk from good to weak' and the 'quality of risk management also falls to acceptable'. These actions finally acted as a trigger form more senior involvement. However, by this time Rusnak had halted his price manipulation and so when back-office staff checked the values they tallied with the external sources.

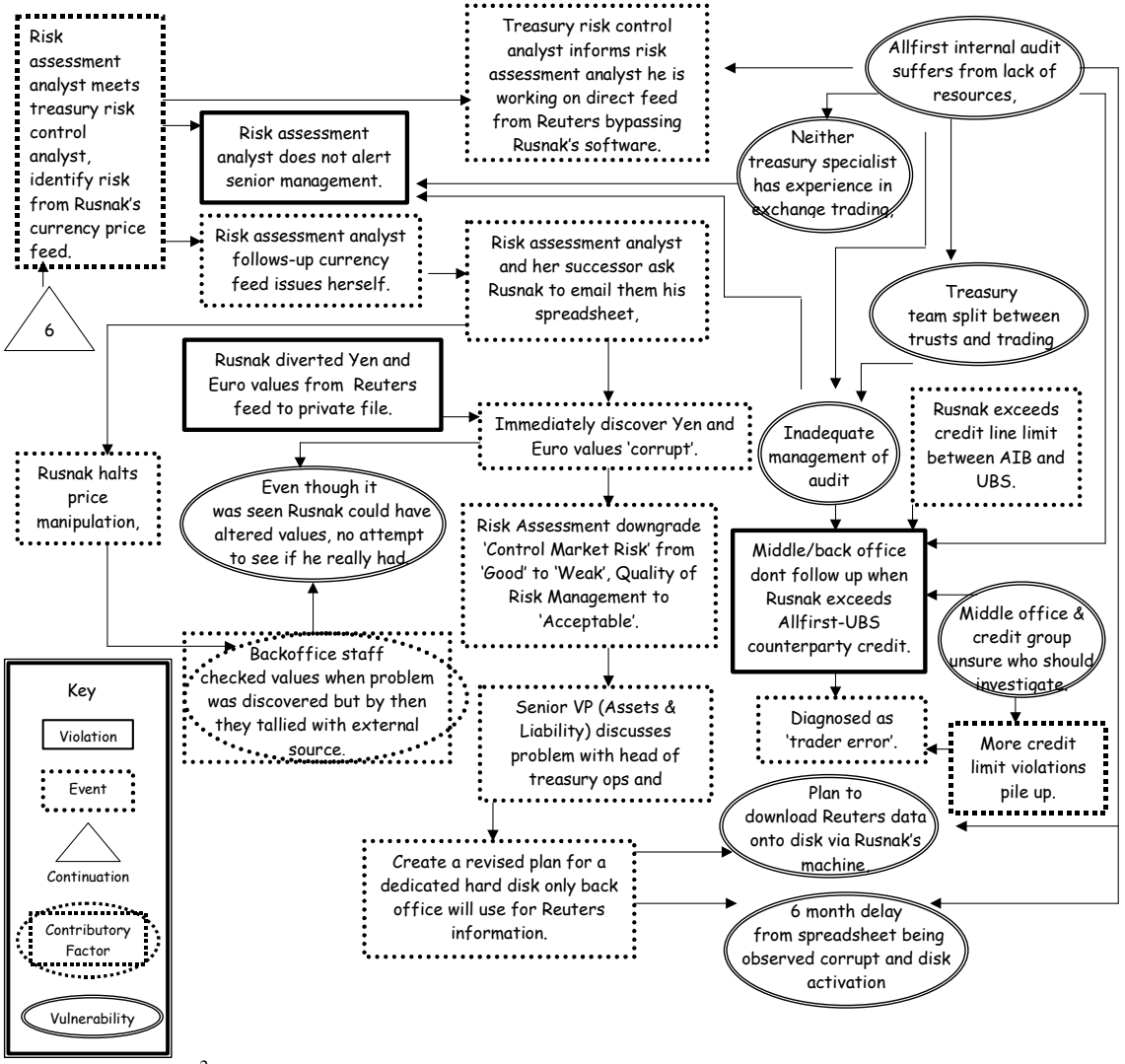


Figure 11: V^2 Diagram Showing Problems in Responding to Reports of Control and Risk Issues

Figure 11 also shows further consequences of the resource constraints imposed on the Allfirst internal audit. The division of one audit team between trust and trading together with problems in the management of these diverse activities led to inadequate oversight for the audit process. At the same time as senior management were becoming aware of the currency feed problems, Rusnak was also exceeding his credit line limit between AIB and UBS. These audit problems partly explain the failure of middle and back-office staff to follow up the reasons for Rusnak exceeding the credit limits. The middle office and credit groups were unsure about who should investigate these problems and in this confusion more credit violations continued to 'pile up'. The lack of thorough audit and the failure to follow-up on these violations partly explains why they continued to be 'diagnosed as trader error' rather than as symptoms of a security violation.

Figure 12 goes on to show the events and contributory factors that led to the discovery of Rusnak's activities. It is important to study this process of discovery. Previous sections have argued that we are unlikely ever to be able to eliminate potential vulnerabilities in security-critical systems. It, therefore, follows that we must learn as much as possible about those defenses that eventually lead to the detection of particular violations. In this case, there is a link between the V^2 diagram and the previous Figure 10 through the continuation symbol labeled 7. The earlier diagram showed that Rusnak had continued to sell year-long 'deep in the money' options. These activities trigger a report from a market source to AIB's Chief Executive that Allfirst is involved in heavy foreign exchange trading. As can be seen at the top of Figure 12, the Allfirst Treasurer responds that there have been no unusual transactions after asking for daily reports on the Allfirst daily foreign exchange transactions. The memo from the AIB Chief Executive was not passed to other senior managers in that bank. After the Treasurer's response from Allfirst, the matter is dropped.

The V^2 diagram in Figure 12 illustrates a further way in which Rusnak's activities might have been discovered. At the end of the 2000 financial year, Allfirst were required to prepare a variety of financial statements. The Allfirst internal audit group questioned the head of treasury funds management on whether Rusnak's use of the balance sheet was justified by the profits that he was able to generate. AIB group's financial reporting unit raised similar questions. As we have seen before, many in the Allfirst senior management were strongly supportive of Rusnak's trading strategy. The explanation that this was assumed to be low-risk together with the lack of any additional questions from fellow traders and the lack of any systematic review of the previous reports in Figure 11 about poor control strategies all contributed to the internal audit decision to drop their investigations. Similarly, the Allfirst controller, director of finance and head of treasury all meet to allay the concerns raised by the AIB financial reporting unit.

Rusnak's continued options trading were eventually mentioned in a letter to Allfirst from the Security and Exchange Commission. The Allfirst financial reporting unit found that the large offsetting positions created by Rusnak were a potential source of risk. At the same time, AIB requested a report on Allfirst's activities for the Central Bank of Ireland. AIB then learn of the increasing foreign exchange transactions and call the Allfirst treasurer. The treasurer then ordered a further investigation. This elicits the response shown as a violation in Figure 12 'Rusnak argues the reports are incorrect using trade dates and not year end values'. Again this line of investigation seems to falter. However, together with the lines of enquiry mentioned above, it does form part of a growing suspicion about the trader's activities.

The final detection factor in this V^2 diagram is prompted by the discovery of unconfirmed exchange tickets by back-office staff. Normally exchange options are marked on tickets that are then annotated to indicate that they have been successfully confirmed as 'legitimate' with the named counterparties. The supervisor who noticed these tickets then asked their staff to gain confirmation, which had not been usual practice for Rusnak's trades as explained in Figure 8. The supervisor is eventually told that the trades with Asian counterparties are bogus. Meanwhile as a result of the Allfirst Treasurer's previous request for daily reports on exchange transactions, he notices a spike in exchange trading that can be linked to Rusnak's activities. He, therefore, proposed to Rusnak's supervisor that his positions be closed. These two lines of investigation combine in the continuation symbol 8 that provides a link with the subsequent V^2 diagram in Figure 13.

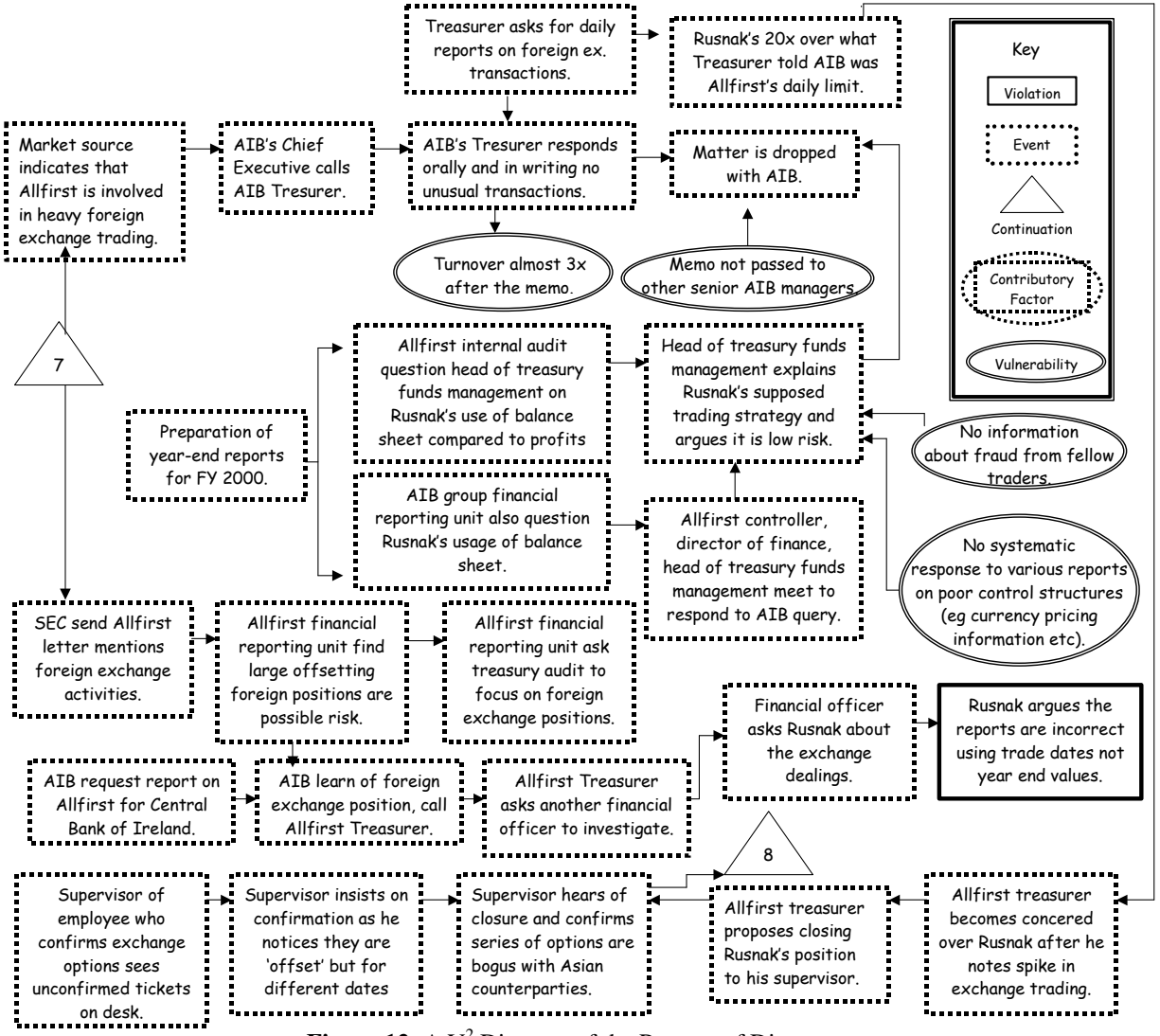


Figure 12: A V^2 Diagram of the Process of Discovery

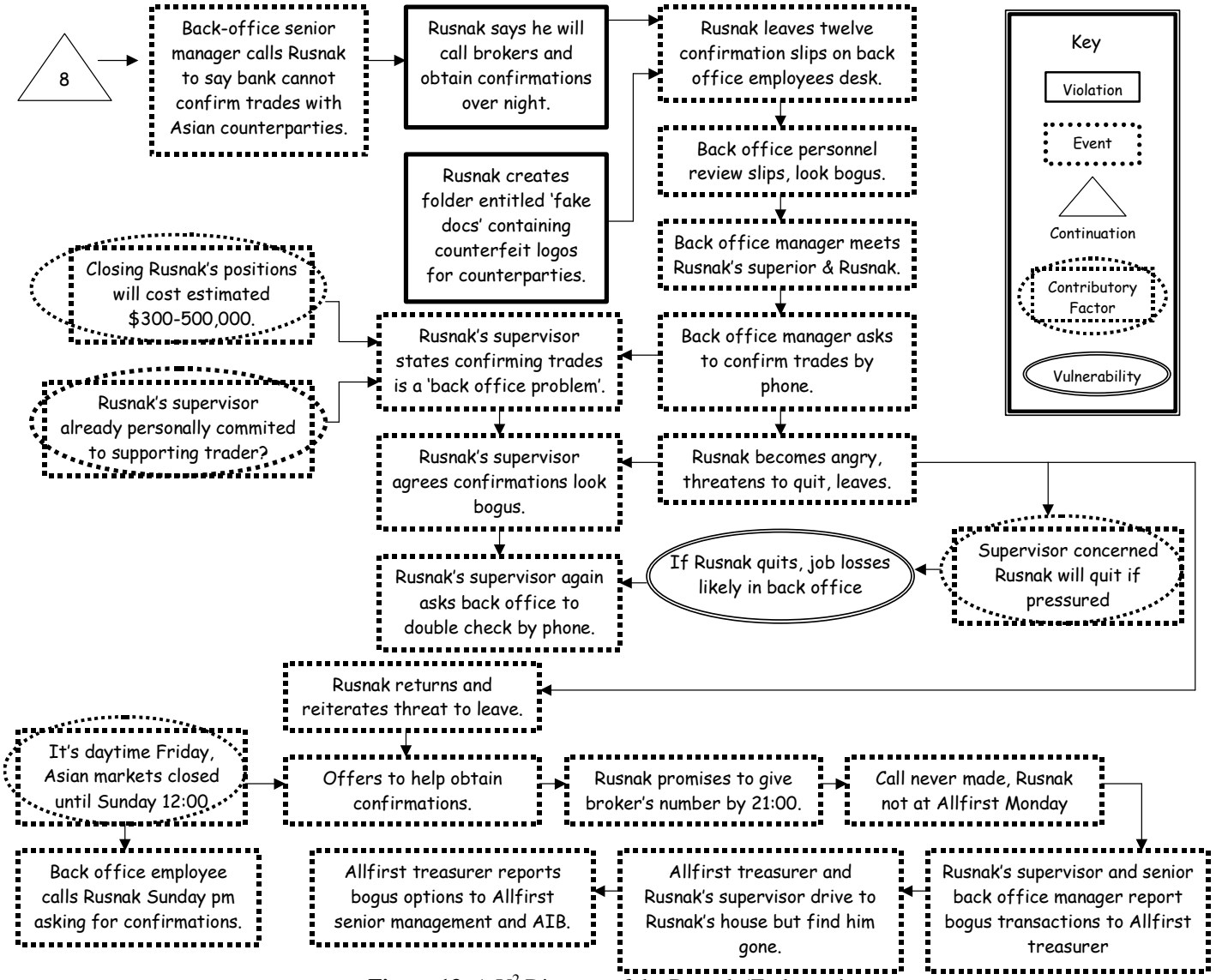


Figure 13: A V^2 Diagram of the Rusnak 'Endgame'

Figure 12 illustrates the start of the discovery process. One of the back-office supervisors finds Rusnack's unconfirmed option tickets and discovers that they denote bogus trades. At the same time, the Allfirst treasurer becomes aware of spikes in the bank's foreign exchange trading that he thought had been brought within tight limits. Figure 13 continues the analysis. A can be seen, the supervisor's senior manager called Rusnack to notify him that they cannot confirm the trades with the counterparties. Rusnack delays the investigation by a violation labeled 'Rusnack says he will call the brokers to obtain confirmations over night'. At the same time, he created a folder on his personal computer entitled 'fake docs'. This was subsequently found to contain counterfeit logos and other information relating to the supposed counterparties for the various option transactions.

The V^2 diagram goes on to show the events that led from Rusnak's delivery of twelve apparently 'confirmed' option slips to the back-office. The back-office manager believed them to have been forged and so decided to consult with both Rusnak and his superior. The back-office manager argued that the trades should be confirmed by telephone at which point Rusnak became angry and threatened to quit. It is important to reiterate that these events are just as relevant to an investigation into a security violation as the technical and managerial vulnerabilities that created the opportunity for the fraud. As we have seen, previous warnings had been overlooked or ignored. Even at this relatively late stage, it might have been possible for many aspects of the fraud to go undetected. For instance, Figure 13 denotes that Rusnak's supervisor was concerned that he would quit if he were pressurized too much about his options trading. These concerns partly stemmed from the fact that back-office jobs would be threatened if his trader resigned. These concerns represent a potential vulnerability that could have persuaded the middle management to ignore the warnings they had received about Rusnak's activities. Rusknak's supervisor also argued that confirming trades was a back-office problem. Again, this response may have been motivated by the estimated $300,000-$500,000 that it would cost to close his positions. It may also have been motivated by the personal support that the manager had provided for his traders supposed activities in previous years. Rusnak's supervisor agrees that the confirmations looked bogus but asked the back-office staff to again seek confirmation over the phone.

Rusnak later returned to the meeting between the back-office manager and his supervisor. He offered help to confirm the transactions. However, it is Friday and the Asian markets will be closed until Sunday midday. Rusnak promises to give them the broker's telephone numbers by 21:00. The call is never made. A back-office employee rings Rusnak on Sunday afternoon asking for the confirmations and their associated telephone numbers but cannot reach Rusnak. Rusnak does not appear at his desk the following Monday. His supervisor and the senior back-office manager then report the bogus transactions to the Allfirst treasurer. The treasurer joined Rusnak's supervisor in driving to the trader's house but they find that he has left. The Allfirst treasurer then passes his concerns on to others in the senior management of Allfirst and of the AIB group.

The previous pages have shown the way in which V^2 diagrams can be used to map out the events and contributory factors, the violations and the vulnerabilities that characterize serious security incidents. The intention has been to provide a detailed case study so that this approach might be extended to other adverse events. This approach also helps investigators to focus on the detection factors that combine to help organizations identify that they may have a potential problem. In Rusnak's fraud there were several opportunities where his violations might have been exposed. These range from external reports, such as market sources questioning the extent of foreign exchange dealing at Allfirst through to regulatory intervention, such as the questions asked in response to the report required by the Irish Central Bank. Staff vigilance also played a role. Even though the Allfirst internal audit teams were ill-prepared to identify Rusnak's actions they did notice problems in the currency feed. As we have seen, however, the V^2 diagrams map out the various factors that combined to divert or extinguish these lines of enquiry. Key personnel had significant investments, in terms of time and reputation, in the success of Rusnak's activities. They were also aware that the future of their own careers and those of their colleagues depended to some extent on the trader's operations. At other times, several members of staff decided to take personal responsibility for investigating their concerns rather than asking more senior management to conduct a more sustained enquiry. Finally and above all, the links between audit and risk management were never clearly established. Doubts about the accuracy of the key VaR metric and about the security of the currency feeds never triggered the sustained audit that might have disclosed the fraud at a relatively early stage.

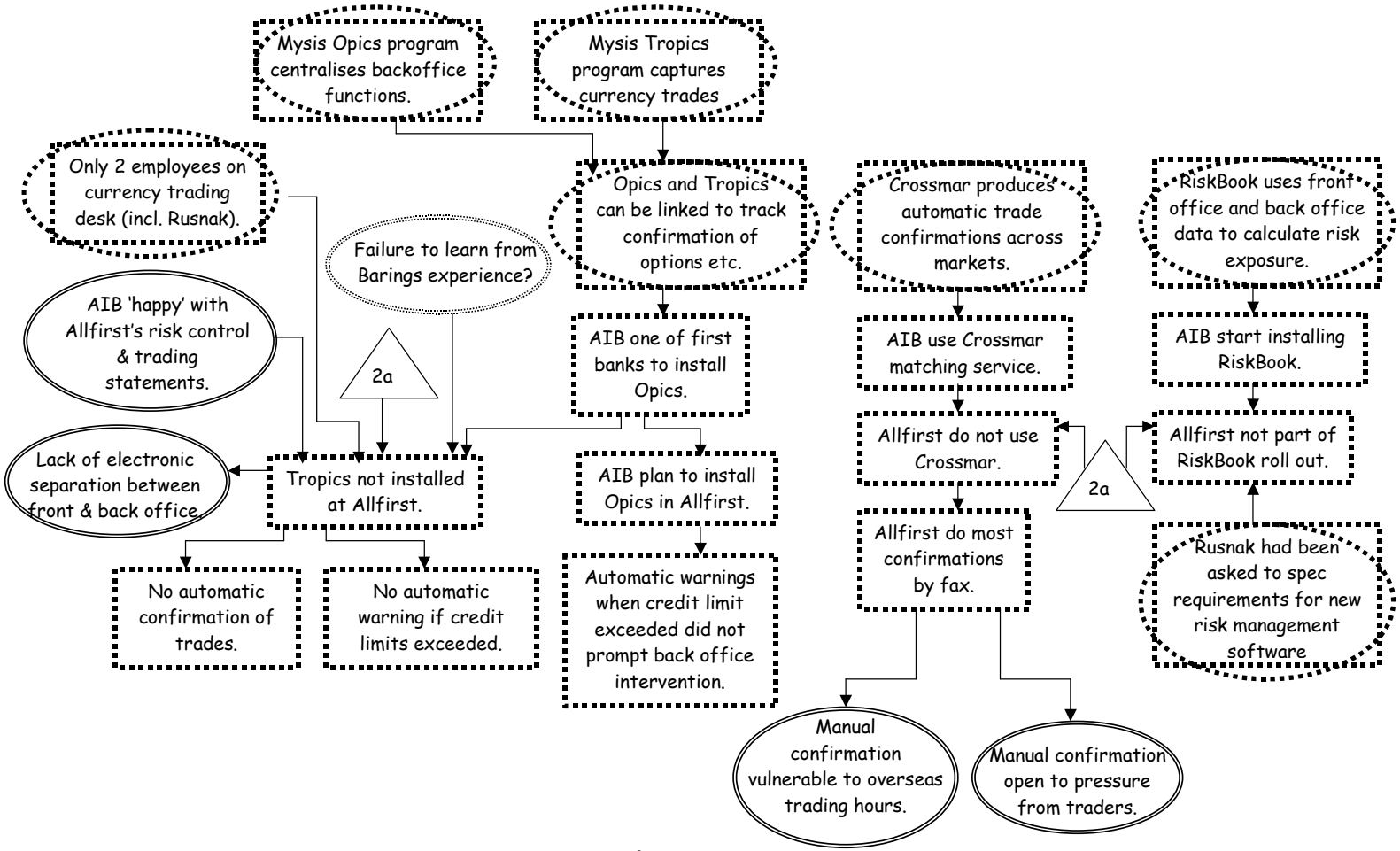


Figure 14: A V^2 Diagram of Software Issues

The previous V^2 diagrams have focused on the construction of an event-based model of the Rusnak fraud. There are other ways in which this technique can be used. Diagrams can also focus in on particular aspects of a security related incident. For example, Figure 14 shows how a V^2 diagram can be constructed to look more narrowly at the role that software based systems played in the fraud. This is particularly important given continuing concerns about the management and oversight of access provided by this class of applications. The continuation symbol labeled 2a refers back to Figure 6. This described some of the contextual factors that stemmed from the merger between Allfirst and AIB. In particular, it relates to AIB's decision that Allfirst should be allowed considerable independence and that the new acquisition should be managed with a 'light hand'. AIB had been one of the first banks to invest in a software system called Opics. The Opics application automates and centralizes a number of back-office functions. It can also be used in conjunction with a 'sister-application' known as Tropics that supports currency trading. An important benefit of using these applications together is that they can enforce a separation of back-office and front-office activities. They can also be used to trace the confirmation of options that were created by the front-office staff and should have been monitored by back-office employees. Tropics was not installed at Allfirst. Hence the software did not support the tracking and clear division of responsibilities that might have prevented many of the vulnerabilities and violations that were identified in previous V^2 diagrams.

As can be seen in Figure 14, the decision not to install Tropics was justified on many grounds. Firstly, the costs of the software may not have been justified by the relatively small size of the trading desk. Also, at the time of merger AIB appeared to be happy with the Allfirst risk control and trading statements. They arguably did not see any justification for the additional monitoring facilities provided by the Tropics application. The decision to invest in Tropics can also be partly explained by a failure to learn from the Barings experience where a trader had managed to erode the separation between front and back office functions. Finally, there was no tradition for preserving this separation in terms of the electronic systems that support the work of Allfirst staff. The outcomes from the decision not to install Tropics included the lack of any automatic confirmation for trades. The decision not to install Tropics also prevented any automatic warnings for traders when their activities exceeded credit limits.

Figure 14 illustrates how V^2 diagrams can be used to gradually piece together more detailed information from a variety of sources. These included the official initial investigation (Promontory, 2002) as well as a number of subsequent reports (Gallager 2002, de Fontnouvelle, Rosengren, DeJesus-Rueff and Jordan, 2004). These sources reveal that Allfirst did go ahead with the installation of the Opics back-office modules associated with the Tropics front-office application. This did help to generate warnings when credit limits were exceeded. However, as we have seen from Figure 11, a host of technical and organizational factors persuaded the back-office staff that these warnings indicated numerous trader errors rather than significant alarms about bogus trading activities.

In addition to the Opics and Tropics systems, Allfirst might have been protected by the introduction of the Crossmar software that was used by AIB. This application also provided automated confirmation for trades using a matching service. Allfirst did not use the Crossmar software and so most of the confirmation relied upon back-office staff to fax requests to overseas markets. This manual confirmation was vulnerable to interruption and dislocation due to overseas trading hours. It was also open to pressure from traders such as Rusnak. Although we have not included it in the current analysis, Figure 14 might also be extended to illustrate the additional pressures that Rusnak's activities created for the back-office staff. His bogus options relied upon the continual generation of additional transactions beyond his legitimate trading activity. One side-effect of the fraud would, therefore, have been to increase the workload on back-office staff which in turn may have left them even more vulnerable to attempts to delay or ignore confirmations on a rising number of trades. AIB had also decided to exploit a software application known as RiskBook. This uses front and back-office systems to calculate the bank's risk exposure. Previous sections have described how Rusnak was able to affect the VaR calculations and there is reason to suppose that the use RiskBook might have offered some protection against these actions. Allfirst were not, however, part of the first roll-out for the RiskBook software within Allfirst. It is deeply ironic that Rusnak had been asked to specify the requirements for this new risk management software.

CONCLUSIONS AND FURTHER WORK

A number of commercial and governmental organizations have recently argued that we must look beyond the immediate events that surround security-related incidents if we are to address underlying vulnerabilities (Austin and Darby, 2003). It is important to look beyond the immediate acts of 'rogue traders' or individual employees if we are to correct the technical and managerial flaws that provide the opportunities for security to be compromised. This paper has, therefore, provides an introduction to Violation and Vulnerability analysis using V^2 diagrams. The key components of this technique are deliberately very simple; the intention is to minimize the time taken to learn how to read and construct these figures. The paper has, in contrast, been motivated by a complex case study. The intention has been to provide a sustained example at a level of detail that is appropriate to an initial investigation into complex security incidents. Previous pages have provided a sustained analysis of Rusnak's fraudulent transactions involving the Allfirst bank. This case study is appropriate because it involved many different violations and vulnerabilities. These included failures in the underlying audit and control mechanisms. They included individual violations, including the generation of bogus options. There were also tertiary failures in terms of the investigatory processes that might have uncovered the fraud long before bank personnel eventually detected it.

Much remains to be done. We are currently working with a number of organizations to extend and tailor the techniques in this paper to support security investigations in a range of different fields, including both financial and military systems. There is a common concern that the V^2 approach will provide a standard means of representing and modeling the outputs of an investigation into the causes of security-related incidents. In each case, however, we are being encouraged to extend the range of symbols represented in the diagrams. For example, these might be used to distinguish between different types of barriers that should have led to the identification of a violation or vulnerability. In terms of the Allfirst case study, the decision not to tell senior management about concerns over the Reuter's currency feed via Rusnak's PC would have to be represented using a different type of symbol. The intention is that analysts would then be encouraged to probe more deeply into the reasons why this potential warning was not acted upon. An important concern in this continuing work is, however, that the additional notational elements will increase the complexity of what is a deliberately simple approach. It is critical to avoid additional complexity in the analysis of what are almost always extremely complex events.

Further work also intends to explore the use of V^2 diagrams as a communication tool with wider applications. In particular, the outcomes of many security investigations must be communicated to diverse groups of stakeholders. These are not simply confined to security professionals and senior management in the target applications. In particular, it is often necessary to communicate findings about the course of an incident with members of the public who may potentially be called upon to act as jurors in subsequent litigation. The complexity of many recent security related incidents makes it vitally important that we find the means to help people understand the events and contributory factors that form the context for many adverse events. Similarly, political intervention is often triggered by incidents such as the Allfirst fraud. It can be difficult to draft effective legislation when key figures lack the necessary time and briefing material to fully follow the events that they seek to prevent.

REFERENCES

R.D. Austin and C.A.R. Darby, The Myth of Secure Computing, Harvard Business Review, (81)6:120-126, 2003.

BBC News, Bank sues over $700m fraud, British Broadcasting Company, London, BBC On-Line, 23 May 2003.

Cisco, Network Security Policy: Best Practices White Paper, Technical report number 13601, Cisco Systems Inc., San Jose, USA, 2003.

US Department of Energy, Root Cause Analysis Guidance Document, Office of Nuclear Safety Policy and Standards, Guide DOE-NE-STD-1004-92, Washington DC, 1992.

US Department of Energy, DOE Standard Safeguard and Security Functional Area, DOE Defense Nuclear Facilities Technical Personnel, Standard DOE–STD–1171–2003, Washington DC, 2003.

P. de Fontnouvelle, E. Rosengren, V. DeJesus-Rueff, J. Jordan, Capital and Risk: New Evidence on Implications of Large Operational Losses, Federal Reserve Bank of Boston, Boston MA, Technical Report, 2004.

S. Gallacher, Allfirst Financial: Out of Control, Baseline: Project Management Information, Ziff Davis Media, March 2002.
G.L. Jones, Nuclear Security: Improvements Needed in DOE's Safeguards and Security Oversight, US General Accounting Office, Washington DC, Report GAO/RCED-00-62, 2000.
C.W. Johnson, A Handbook of Incident and Accident Reporting, Glasgow University Press, Glasgow, Scotland, 2003.
K. Julisch, Clustering intrusion detection alarms to support root cause analysis. ACM Transactions on Information and System Security, (6)4:443–471, 2003
G. Killcrece, K.-P. Kossakowski, R. Ruefle, M. Zajicek, Organizational Models for Computer Security Incident Response Teams (CSIRTs), Technical Report CMU/SEI-2003-HB-001, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, USA, 2003.
J. Lew, Guidance On Implementing the Government Information Security Reform Act, Memorandum for the Heads of Departments and Executive Agencies, Whitehouse Memorandum M-01-08, Washington DC, 2001.
J.L Mackie, (1993), Causation and conditions. In E. Sosa and M. Tooley (eds.), Causation and Conditions, pages 33-56. Oxford University Press, Oxford, 1993.
C.A. Meissner and S.M. Kassin, "He's guilty!": investigator bias in judgments of truth and deception. Law and Human Behavior, 26(5):469-80, 2002.
Microsoft, Microsoft Solutions for Securing Windows 2000 Server, Microsoft Product & Technology Security Center, Redmond USA, 2003. Available from http://www.microsoft.com/technet/security/prodtech/win2000/secwin2k/default.mspx
Naval Surface Warfare Centre, Dahlgren, Computer Security Incident Handling Guidelines, Department of the Navy, Commanding Officer, Fleet Information Warfare Center, Virginia, USA, 2002.
T. Oberlechner, The Psychology of the Foreign Exchange Market, John Wiley and Sons, New York, USA, 2004.
Promontory Financial Group, Report to the Board and Directors of Allied Irish Bank PLC, Allfirst Financial Inc. and Allfirst Bank Concerning Currency Trading Losses Submitted by Promontory Financial Group and Wachtell, Lipton, Rosen and Katz, First published by Allied Irish Banks PLC, Dublin, Ireland, March 2002.
A.M. Rabinowitz, The Causes and Some Possible Cures: Rebuilding Public Confidence in Auditors and Organizational Controls
Certified Public Accountants Journal, 66(1):30-34 1996.
J. Reason. Managing the Risks of Organizational Accidents. Ashgate Publishing, Aldershot, 1997.
K. Roark, Los Alamos Director Testifies on Security Incident, Lawrence Livermore National Laboratory, Press Release, Livermore, California, USA, June 2000.
S. Skalak, Financial Fraud: Understanding the Root Causes, Price Waterhouse Cooper, Financial Advisory Services,Dispute Analysis & Investigations Department (2003).
P. Stephenson, Modeling of Post-Incident Root Cause Analysis, International Journal of Digital Evidence, (2)2:1-16, 2003.
L. Tvede, The Psychology of Finance, John Wiley and Sons, New York, 1999
M.J. West-Brown, D. Stikvoort, K.-P. Kossakowski, G. Killcrece, R. Ruefle, M. Zajicek, Handbook for Computer Security Incident Response Teams (CSIRTs), Technical Report CMU/SEI-2003-HB-002, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, USA, 2003.