MASTER

**An
Aggregated Systems Model
For
Nuclear Safeguards Decisions**

March 1979

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

009897503

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Making the United States or the United States Government of Training, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

CONTENTS

	Page
LIST OF FIGURES.	v
PREFACE.	ix
I EXECUTIVE SUMMARY	1
Introduction.	1
Decision Environment	1
Advantages and Limitations of the Analytical Tool.	2
Example of Specific Safeguards Decisions	3
Aggregated Systems Model.	5
Overview	5
ASM Capabilities and Uses.	5
Description and Illustrative Application of the ASM	9
Facility Model	9
Adversary Model.	11
Interaction: Facility and Adversary Models.	13
Illustrative Public Consequence Model.	16
Utility Model.	18
Sample Results	18
Summary	28
II AGGREGATED SYSTEMS MODEL DESCRIPTION.	29
Introduction.	29
Background	29
Definition	30
Relation to Other Work	30
Report Organization.	30
Aggregated Systems Model Overview	31
Overview Figure	31
Using the ASM.	31
Differences from an Earlier Aggregated Systems Model.	34
Assumptions in the Analysis.	36
Facility Model.	37
Adversary Model	41
Adversary Types.	41
Diversion Strategy and Tactics	41
Adversary Utility Model.	43

	Page
II AGGREGATED SYSTEMS MODEL DESCRIPTION (Continued)	
Adversary-Facility Interaction.	49
Schematic Decision Tree Description.	49
Extensive Form of the Decision Tree.	54
Markov Model	58
Consequence Model and Social Utility Model.	65
III ILLUSTRATIVE DATA	71
Purpose	71
ASM Influence Diagram--Probabilistic Dependence	71
Important Dependence Assumptions	71
Detection: Aggregating Component Performance	73
Safeguard System Designs.	76
Adversaries	76
Diversion Strategies.	76
Detection	79
Diversion Tactics	88
Identification.	88
Interruption.	88
Adversary Utilities	88
IV ILLUSTRATIVE ANALYSIS	95
Introduction.	95
Adversary Decision Analysis	96
Adversary Tactical Decisions	96
Adversary Strategic Decisions--General	
Description.	99
Adversary 4's Strategies	106
Adversaries 9, 11, and 12 Strategies	109
Safeguards Design Evaluation.	111
Overview	111
Design Sensitivity	116
Marginal Component Benefit	116
Performance Graphs.	119
Adversary Sensitivity	124
Summary	124
V CONCLUSIONS	127
APPENDICES	
A. ADDITIONAL INFORMATION ON ADVERSARY TACTICS	A-1
B. MATHEMATICAL DERIVATION OF RESULTS	B-1

GLOSSARY.	G-1
REFERENCES.	R-1

LIST OF FIGURES

<u>Chapter I</u>	Page
1.1 Aggregated System Model Overview.	6
1.2 Components of Safeguards System Designs	10
1.3 Adversary Characteristics	12
1.4 Decision Tree: Adversary - Facility Interaction.	15
1.5 Sample Strategic Choice by Adversary 4 Given Design 2 . . .	17
1.6 Illustrative Consequence Model.	19
1.7 Comparison of Adversary Strategy and Tactics for Two Designs	20
1.8 Evaluating the Two-Person Rule.	22
1.9 Comparison of Frequent Inventory and Quantity Estimation Monitors.	23
1.10 Performance Graph: Cost Versus Expected SNM Diverted Per Year.	24
1.11 Value/Impact Tradeoff Curve: Safeguards Cost Versus Diversion Cost.	26
 <u>Chapter II</u>	
2.1 Aggregated Systems Model Overview	32
2.2 Elements of a Safeguards System Design.	38
2.3 Adversary Characteristics	42
2.4 Adversary Utility Function.	45
2.5 Assessment Lotteries for the Adversary Utility Model. . .	46
2.6 Decision Tree: Adversary - Facility Interaction.	50
2.7 Repeated Attempt Sequence Events (Schematic Tree)	52
2.8 Relation Between Outcome States and Attempt Sequence Events.	55

	Page
<u>Chapter II (Continued)</u>	
2.9 Decision Tree For Each Try	57
2.10 Repeating Lottery on the Outcome States From Each Try. .	59
2.11 Markov Model	60
2.12 Decision Tree Representing Final Outcome From Repeated Trials	62
2.13 Equations For Computing Markov Statistics.	63
2.14 Illustrative Consequence Model	66
2.15 Consequence Model Probabilities.	67
2.16 Expected Consequences by Adversary and Quantity Diverted (Equivalent $\$10^6$)	69

Chapter III

3.1 Diversion Model Influence Diagram.	72
3.2 Influence Diagram with Detection System Detail	74
3.3 Safeguards System Designs.	77
3.4 Adversary Probabilities for Acquiring SNM.	78
3.5 Diversion Strategies	80
3.5(2) Diversion Strategies (Continued)	81
3.6 Probability Tree for Detection Assessment.	82
3.7 Influence Diagram for Detection Probability Assessment .	84
3.8 Detection Probability Assessments.	85
3.8(2) Detection Probability Assessments.	86
3.8(3) Detection Probability Assessments.	87
3.9 Identification Probability Assessment.	89
3.10 Interruption Probabilities	90

	Page
<u>Chapter III (Continued)</u>	
3.11 Adversary Utility Function Assessment Form	91
3.12 Adversary Utility Function Parameters.	92
3.13 Forms of Selected Adversary Utility Functions.	93
<u>Chapter IV</u>	
4.1a Example Evaluation of Tactics -- Outsider.	97
4.1b Example Evaluation of Tactics -- Insider	97
4.2 Adversary Tactics.	98
4.3 Adversary 1's Tactical Decision Problem.	100
4.4 Adversary 3's Strategic Decision Evaluation.	101
4.5 Adversary 3's Choice of Tactics in Diversion Strategy 2	103
4.6 Evaluation of All Adversaries' Strategies -- Design 1. .	105
4.7 Evaluation of All Adversaries' Strategies -- Design 2. .	107
4.8 Sample Strategic Choice by Adversary 4 Given Design 2. .	108
4.9(1) Adversary Decisions for All System Designs	112
4.9(2) Adversary Decisions for All System Designs	113
4.9(3) Adversary Decisions for All System Designs	114
4.10 Summary Performance Measures for All Designs	115
4.11 Marginal Evaluation of Safeguards Components	117
4.12 Performance Graph: Cost Versus Probability of Detection.	120
4.13 Performance Graph: Cost Versus Expected SNM Diverted Per Year	121
4.14 Value/Impact Tradeoff Curve: Safeguards Cost Versus Diversion Cost	122

	Page
<u>Chapter IV (Continued)</u>	
4.15 Adversary Decisions -- "Fearless" Adversary Sensitivity	125
4.16 Summary Performance Measures -- "Fearless" Adversary Sensitivity	126

Appendix

A1 Diversion Tactics Analysis.	A-2
A2 Possible Endpoints in Figure A1	A-4
A3 Actual and Perceived Probability Trees.	A-6
A4 Markov Model With Late Identification	A-7

PREFACE

This report summarizes a general analytical tool designed to assist nuclear safeguards decision-makers. The approach is based on decision analysis--a quantitative procedure for evaluating complex decision alternatives with uncertain outcomes.

The report describes the general analytical approach in the context of safeguards decisions at a hypothetical nuclear fuel reprocessing plant.

The approach and the existing model were developed by Applied Decision Analysis (ADA) and Lawrence Livermore Laboratory (LLL) for the Nuclear Regulatory Commission. We wish to acknowledge the support of Dr. Stein Weissenberger, Dr. Ivan Sacks, and Dr. Rokaya Al-ayat, of LLL, all of whom contributed in a substantial way to the analysis. The authors are Dr. Bruce R. Judd and Jean M. Huntsman of ADA. The contract is University of California Purchase Order 9897503 with Applied Decision Analysis, Inc.

I. EXECUTIVE SUMMARY

Introduction

Decision Environment

This report describes an analytical tool for people who make nuclear safeguards decisions. This tool is designed to help the Nuclear Regulatory Commission (NRC) set performance criteria for safeguards systems in facilities handling special nuclear material (SNM). It will also help facility designers choose safeguards systems to meet NRC's criteria.

In addition, the tool is useful for evaluating security at a given facility. It provides a "first-cut" assessment of safeguards system performance, the results of which can guide more detailed assessment. Finally, when the detailed assessment is concluded, this tool can be used to summarize the results.

We address four factors that complicate safeguards decision-making:

- Uncertainty. The adversary's motives, resources, knowledge, and plan of attack are all uncertain. Because safeguards systems are not perfectly reliable or immune from tampering, their ability to thwart the attempt is also uncertain. In addition, the consequences of diverted SNM cannot be estimated precisely.
- Data. When setting criteria, the regulator must rely on a mix of objective technical data and subjective judgment. Safeguards performance characteristics can be tested to some extent; on the other hand, adversary characteristics and the consequences of diverted SNM are still in the domain of expert opinion.
- Balancing risk and cost. When setting criteria, the regulator must choose the point at which the social cost of additional safeguards does not exceed the social benefit.

I. EXECUTIVE SUMMARY

Introduction

Decision Environment

This report describes an analytical tool for people who make nuclear safeguards decisions. This tool is designed to help the Nuclear Regulatory Commission (NRC) set performance criteria for safeguards systems in facilities handling special nuclear material (SNM). It will also help facility designers choose safeguards systems to meet NRC's criteria.

In addition, the tool is useful for evaluating security at a given facility. It provides a "first-cut" assessment of safeguards system performance, the results of which can guide more detailed assessment. Finally, when the detailed assessment is concluded, this tool can be used to summarize the results.

We address four factors that complicate safeguards decision-making:

- Uncertainty. The adversary's motives, resources, knowledge, and plan of attack are all uncertain. Because safeguards systems are not perfectly reliable or immune from tampering, their ability to thwart the attempt is also uncertain. In addition, the consequences of diverted SNM cannot be estimated precisely.
- Data. When setting criteria, the regulator must rely on a mix of objective technical data and subjective judgment. Safeguards performance characteristics can be tested to some extent; on the other hand, adversary characteristics and the consequences of diverted SNM are still in the domain of expert opinion.
- Balancing risk and cost. When setting criteria, the regulator must choose the point at which the social cost of additional safeguards does not exceed the social benefit.

- Communications. Information must flow from scientists and engineers to regulators, to law makers, and to the public. Safeguarding SNM is a technological problem, involving highly specialized experts in fields from law enforcement to physics. Simultaneously, it is a social problem, incorporating the public's perceptions and values.

Advantages and Limitations of the Analytical Tool

The analytical tool described here helps the decision-maker who confronts these factors:

- Uncertainty. The tool considers uncertainty explicitly. Using a dynamic and probabilistic model, the tool helps the decision-maker investigate the system's vulnerability to many types of adversaries, some of whom may make repeated covert attempts to divert SNM.
- Data. We will never have precise data on all adversaries, system performance, and consequences. Because judgmental input is required, the tool does not yield exact answers to these safeguards questions. However, it does make use of existing performance data, where it is available. Further, it shows which judgmental data have the greatest influence on the safeguards decisions, and where additional research is needed.
- Balancing risk and cost. The tool improves the consistency of safeguards criteria by making explicit and balancing quantitative judgments on both sides of the benefit-cost tradeoff. Through sensitivity analysis, it identifies the value judgments that are most crucial to the criteria-setting decision.
- Communications. The tool is designed to assist communication among parties in the decision process. Although it requires some technical inputs, it is sufficiently aggregated and simple so that non-technical people can understand the link between technical inputs and model outputs. Thus it can play an important role in the process of setting and justifying regulations. It

documents major assumptions and technical judgments behind the decision and allows them to be "checked." As information changes, the tool can be used to determine the appropriate change in safeguards criteria, if any.

Although it has many benefits, the tool has limitations. First, the models we have developed are highly aggregated and thus they do not represent the same degree of information found in detailed facility assessments. Second, the tool is not a "computer formula" for making decisions; instead, it aids the human decision-makers, reflecting their judgment, not replacing it. Third, the tool is only as useful as the information that feeds it, and unfortunately, the current data base is "sparse." Much of the current data is judgmental. Fourth, explicitness, which is generally beneficial, may be deleterious in some instances. With data from an actual facility (not included in this report), the model and especially its input calculations could show the adversary the path to successful diversion. Furthermore, some people question the wisdom of making explicit social value judgments in a political environment. We believe, however, that the consistent application of value judgments, the insight gained through explicit analysis, and the logical justification of regulations is a benefit that exceeds the political risk.

Example of Specific Safeguards Decisions

The analytical tool is useful for making safeguards decisions at several different levels. For instance, it can help set performance requirements for individual components, for subsystems such as material control or physical security, or for the entire safeguards system at a facility. In addition, it can help a designer choose which components to include or exclude, or evaluate the tradeoffs when substituting one component for another.

We illustrate the spectrum of uses for setting performance criteria for a nuclear reprocessing plant. We will show the results from analyzing three sample criteria decisions. This type of analysis for setting criteria is often called value/impact analysis. We will also show value/impact analysis results for a large number of facility designs. The three sample decisions will:

- Require a two-person rule for all plant operations
- Substitute for frequent physical inventory a real-time SNM quantity estimator
- Specify the level of performance required from the overall facility safeguards system -- for example, a limit on the expected quantity of SNM diverted per year

The analytical tool is also useful when assessing whether or not a given facility meets performance criteria. Although we do not give such an example in this executive summary, these additional uses are described in the main report.

Aggregated Systems Model

Overview

The tool we have developed to analyze these decisions is called the Aggregated Systems Model (ASM): "Aggregated" because it is a combination of highly aggregated models, and "Systems" because it takes a broad perspective, representing many of the major factors in setting safeguards criteria.

The ASM is a collection of models shown in Figure 1.1. The collection includes:

- Adversary Model. A probabilistic model of several types of adversaries, their attempt frequency, their resources, and their plan of attack (called diversion strategy), including how many tries it will take before they have stolen the desired quantity of SNM.
- Facility Model. This describes the likelihood that each adversary will be detected and interrupted by the safeguards system.
- Safeguards Technology Model. This element describes safeguards components that could be incorporated in the safeguards design. Ultimately, data for the Facility Model and Safeguards Technology Model will be supplied by detailed analyses of facilities.
- Consequence Model. The Consequence Model is a probabilistic statement of the malevolent uses and effects of diverted SNM.
- Utility Model. Outcomes such as diversion consequences and safeguards costs are expressed in common units using the Utility Model. It allows the decision-maker to make explicit tradeoffs among lives lost, property damage, evacuation cost, which are called collectively diversion costs, (C_D), and the cost of safeguards (C_S).

ASM Capabilities and Uses

We can use Figure 1.1 to conceptualize how the ASM will help analyze the types of safeguards decisions mentioned earlier. System design or

Figure 1.1 Aggregated Systems Model Overview

operating procedure decisions such as the two-person rule or substituting quantity estimates for inventories require at least the Facility and Safeguards Technology Models. These will produce measures of facility performance against a specified adversary threat, and they will show how performance changes as the design or operating procedures are varied. When the Adversary Model is coupled with the Facility and Technology models, one can see how the facility performs against the full spectrum of adversaries, all of whom have optimized their plan of attack against the facility. The same three models are used to determine the overall facility performance as measured, for example, by the expected quantity of SNM diverted each year.

While the models discussed so far help measure system performance, alone they do not indicate the optimum level of performance. In other words, they can measure performance against a criterion, but they do not show the best level for the criterion. Setting the criterion requires a comparison of the cost (impact) and reduced risk (value) for different safeguards criteria. This analytical procedure is often called value/impact analysis.

Value/impact analyses have three steps. The first step is analyzing how the designs or requirements affect the facility's performance against adversaries. In Figure 1.1, this analysis is accomplished in the box labeled Facility Model. Performance can be measured by the arrow labeled "SNM Diverted," which emanates from the Facility Model. If the decision-maker can balance cost (impact) with the benefits of reduced SNM diversion (value) in his or her head, then the value/impact analysis stops here. However, if the decision-maker feels additional quantitative analysis would be helpful, then the analytical process continues with step two.

The second step is to express the public consequence of the SNM diverted. For example, the ASM can describe how a tightening of security reduces the amount of SNM diverted and hence reduces the potential number of deaths or damage due to malevolent use of SNM. This translation occurs in the Consequence Model box shown in Figure 1.1. Once again, the process can stop, and at this point the decision-maker implicitly balances

the cost of safeguards with reduced consequences. However, step three can aid this process.

The third step involves balancing changes in public consequences (values) with the social costs (impacts) of the safeguards to achieve the changed consequences. The Utility Model uses explicit value judgments in this process. It applies these consistently across all consequences. It also shows the sensitivity of various conclusions to changes in important value judgments.

As an aid to detailed assessments, the ASM can use some or all of the models in Figure 1.1. In preliminary assessments, the decision-maker needs at least the Safeguards Technology and Facility Models, especially to evaluate performance against a specified adversary threat. The Adversary Model is used to consider the entire range of threats, rather than only one. The output of a preliminary assessment is a statement of the given facility's performance measures such as probabilities of detection, interruption, or expected SNM diverted.

When the ASM is used to summarize detailed assessments, more of the models may be used. If the goal is to show which facility components are the weakest links in safeguards, only the Facility and Adversary Models are needed. At the other extreme, a statement of the total social risk posed by the plant may require all of the models in Figure 1.1.

Description and Illustrative Application of the ASM

This section describes the models and interactions illustrated in Figure 1.1: (1) Facility Model, (2) Adversary Model, (3) Interaction of Facility and Adversary Models, (4) Consequence Model, and (5) Utility Model. These descriptions are followed by illustrative model results, analyzing three specific decisions. The purpose of this section is to convey a general understanding of the kind of input data required and the types of output available.

The data for the example analyses are purely illustrative. They were developed by a few members of the ADA-LLL project team. Clearly, they should not be regarded as accurate.

Facility Model

The Facility Model represents the system designed to safeguard SNM. The current version of ASM includes nine basic building blocks of this system, called "components." These are listed in the left-hand column in Figure 1.2. (Only five of the 14 designs considered in the main report are shown.)

When referring to the safeguards system "design," we mean the list of components installed in the facility. Changing the design means adding, deleting, or substituting components. Five sample designs are also shown in Figure 1.2.

The performance of individual components determines the performance of the safeguards system. Calculating system performance requires: (1) input data on component performance and (2) an algorithm for aggregating component performance into safeguards system performance. The data and algorithm are discussed in Chapter 3. One step in the algorithm is to aggregate component performance to an intermediate level called "subsystem" performance: subsystems are collections of components that have similar detection characteristics. For instance, in the current version of the ASM, we aggregate components into three subsystems:

- Material control (electronic detection)
- Physical security (visual detection and apprehension)

SAMPLE SAFEGUARDS DESIGNS

COMPONENTS	4	2	5	11	14	SUBSYSTEMS
	Base Case	Minimal Safeguards	Two-Person Rule	Quantity Estimators	Frequent Inventory	
1.1 Quantity Estimators				✓		Material Control
1.2 Process State Monitors	✓		✓	✓	✓	
1.3 Personnel Monitors	✓		✓	✓	✓	
1.4 Procedure Monitors						
2.1 Stationary Guards	✓	✓	✓	✓	✓	Physical Security
2.2 Roving Guards						
2.3 Two-Person Rule			✓			
3.1 Nominal Accounting	✓	✓	✓	✓	✓	Accounting
3.2 Frequent Inventory					✓	

Figure 1.2 Components of Safeguards System Designs

- Accounting (records detection).

The aggregation scheme is shown on the right side of Figure 1.2.

Adversary Model

This subsection describes: (1) the types of adversaries currently in the model, (2) the decisions they make about how to divert SNM, and (3) the way they value the possible outcomes from the attempt.

We began the analysis with 14 different adversary types, whose characteristics are listed in Figure 1.3. For the illustrative data set, sensitivity analysis shows that six of these adversaries dominated the other eight in terms of their contribution to the expected quantity of SNM diverted per year. (Percentage contributions are shown on the right of Figure 1.3.) Therefore, only the adversaries with asterisks in Figure 1.3 were retained for this analysis.

For all types of adversaries, we define an "attempt" as the existence of the adversary who wants to divert SNM. The attempt may consist of one or multiple "tries," or, if the adversary is deterred, no tries at all.

Each of the adversary types will optimize his or her plan of attack depending on the type of safeguards in the facility. We call this adversary choice a "strategic" decision, which includes: whether or not it is worth the risk to try to divert SNM, the quantity stolen each try (increment size), the number of tries, and the diversion path (route to and from the SNM). We define the diversion path in terms of which components could detect the diversion attempt.

We also define adversary "tactical" decisions. These involve whether to abort a try if detected by a safeguards system. Presumably, aborting might reduce the chances of being identified, since the safeguards system might assume the detection is a false alarm. If this is the case, then aborting could improve the adversary's long-run chances for success.

We assume the adversary makes these strategic and tactical decisions based on his or her own preferences for the ultimate outcomes from the attempt. These outcomes are:

Number	Access	Resources		Collusion	Desired Quantity	Tries	Attempt Frequency (Att./1000 Yrs.)	Percent of SNM Diverted
		Equipment	Authority					
1*	Outsider	Major	No	No	Bomb	One	2	38
2	"	Minor	"	"	"	"	.02	.003
3*	Insider	Major	"	Yes	"	"	.29	4.7
4*	"	"	"	"	"	Multiple	2.6	33
5	"	"	"	"	Less	One	.03	.049
6	"	"	"	No	Bomb	"	.03	1.1
7	"	"	"	"	"	Multiple	.26	1.2
8	"	"	"	"	Less	One	.03	.006
9*	"	Minor	"	Yes	"	Multiple	13	13
10	"	"	"	No	"	"	1.4	.14
11*	"	Major	Yes	Yes	Bomb	One	.03	1.2
12*	"	"	"	"	"	Multiple	.15	7
13	"	"	"	"	Less	One	.07	.11
14	"	"	"	"	"	Multiple	.11	.18

*These adversary types are dominant

Figure 1.3 Adversary Characteristics

- Complete success
- Partial success
- Failure (no success or status quo)
- Capture.

Preferences for these uncertain outcomes are represented in a utility function. For example, we hypothesize the following utility function for Adversary #4 in Figure 1.3:

<u>Outcome</u>	<u>Utility</u>
Complete success	1.0
Partial success	.01S + .76
No success (or no try)	.7
Capture	0.0

where S is the quantity of SNM diverted, and

$$.3 \text{ kg} \leq S \leq 10.0 \text{ kg}.$$

We assume the adversary evaluates all possible strategies and tactics by considering the likelihood and utility of all outcomes, and then makes the choice with the highest expected utility. We also assume, for this example, that the adversary knows exactly how the safeguards system will respond to an attempt.

Interaction: Facility and Adversary Models

We have modeled the complex interaction between the adversary and the facility safeguards in some detail. Two factors make the interaction complex. First, the adversary will optimize his or her strategy to find the weakness in the safeguards system. As new safeguards are installed to foreclose one diversion path, we assume the adversary will shift to the next "weakest link." The second-best option may be to make no try to divert, in which case the adversary is deterred. Assuming the adversary is not deterred, an adversary may make repeated covert tries, perhaps waiting until the system is vulnerable due to safeguards component failures. A safeguards component that is "slow but sure to detect" (e.g., accounting) may be valuable in preventing diversion by such a repeating adversary.

Figure 1.4 is a schematic decision tree showing major decisions and outcomes in the adversary-facility interaction. Square nodes represent decisions: an "S" in the box denotes decisions made by the facility designer following NRC guidelines in the best interest of society, and an "A" denotes adversary decisions. Circles represent probability nodes, where the decision-maker is uncertain about the outcomes. We show only two branches at each node, although there can be many.

The first decision we consider is the facility design--which safeguards components to choose. Obviously, this decision is made without perfect foresight about which adversaries (if any) will attack and whether the system can detect or interrupt the attempt. Our three example decisions are special cases of this general design decision.

The second node represents the spectrum of adversary threats. The choices of strategy and tactics made by these adversaries is represented in the third node in Figure 1.4. These decisions consider the probabilities of detection, identification, and interruption leading to the outcome states at the right-hand side of the figure. For some adversaries, the detection/identification/interruption sequence of events is repeated for several tries.

Facility safeguards response to the threat is measured in terms of detection, identification, and interruption. "Detection" means the safeguards authority receives an alarm. However, at that point, the signal is undistinguished from a false alarm. "Identification" means the authority knows the alarm is real and knows enough about the adversary so that the guard force can be dispatched to confront the adversary. "Interruption" means that no SNM crosses the plant boundary on a given try and that the adversaries are captured.

The ASM translates input data on component and system detection, identification, and interruption probabilities--plus information on adversary strategy and tactics--into a probability distribution for the

Figure 1.4 Decision Tree: Adversary - Facility Interaction

attempt outcomes shown at the right in the schematic decision tree. An example decision tree showing Adversary 4's strategic choice of SNM increment size is shown in Figure 1.5. The system design is #2 in our illustrative data set--a system with minimal safeguards. Adversary 4 is a highly sophisticated insider with good equipment and in collusion. He or she considers four strategies:

- Steal the desired quantity (10 kg) in 1 kg increments
- Steal it in 300 g increments
- Steal it in one large quantity (10 kg)
- Don't try.

The figure shows that the most likely outcomes in cases where Adversary 4 makes the try are either capture or complete success. These outcomes are valued according to: (1) how much SNM is diverted and (2) the utility functions described earlier. The strategy with the highest utility (.89) also has the highest probability of complete success and the greatest expected quantity diverted. We assume the adversary would choose "one large quantity" for this particular design.

(In contrast, all other designs considered in our illustrative data set had more safeguards, and for half of the designs Adversary 4 was better off not to make an attempt; in other words, for these designs the "No Try" option had the highest expected utility.)

The advantage of the ASN is its ability to test easily the sensitivity of results as in Figure 1.5 to changes in adversary preferences. In this case, we found the adversary's choice of strategy and tactics (including the decision whether or not to try to divert SNM at all) was most sensitive to the utility assigned to the capture and "No try" states, rather than to performance data on safeguards components.

Illustrative Public Consequence Model

Deciding whether the social benefits of a safeguards design are worth the social costs requires a judgment regarding the social consequences of diverted SNM. Usually this judgment is implicit. However,

Figure 1.5 Sample Strategic Choice by Adversary 4 Given Design 2

Consequence modeling can begin to make a meaningful explicit contribution to this decision process. As they become available, consequence model results may be considered in a probability tree framework like Figure 1.6. Uncertain events in the Public Consequence Model are:

- Material recovery
- Intent of the adversary
- Location of weapons
- Evacuation
- Detonation of weapons
- Yield from an explosive weapon.

The probabilities of these events depend on the type of adversary and diverted quantity of SNM. The illustrative consequences of these events are measured in deaths, injuries, evacuation cost, and property damage.

Utility Model

Measuring the consequences in common units is helpful for making implicit or explicit tradeoffs between safeguards cost and diversion risk. In the Public Consequence Model, we evaluate outcomes in terms of public fatalities, damage, dollar loss, and evacuation costs. We have selected dollars as the common numeraire. Illustrative tradeoff values are: \$100,000/injury and \$1,000,000/death. In other words, this judgment implies society is willing to pay 10^6 to prevent the loss of one statistical life and 10^5 to prevent one statistical injury. Evaluation of each outcome in these utility units is given in the right hand column of Figure 1.6.

Sample Results

Before summarizing the results pertaining to the three sample decisions, we highlight the results of the adversary's decision process. We give results in Figure 1.7 for two system designs: (1) Base Case Design and (2) Minimal Safeguards Design. These were defined in Figure 1.2. Notice that all adversaries will attack Design 2, but most are deterred by Design 4. This is because the adversary's chances for

*Utility Function

$$1 \times \text{Deaths} + .1 \times \text{Injuries} + 1 \times \text{Costs} = \text{Utility (Equiv. } \$10^6)$$

Figure 1.6 Illustrative Consequence Model

Design 4				Design 2	
	<u>Adversary</u>	<u>Strategy</u>	<u>Tactics</u>	<u>Strategy</u>	<u>Tactics</u>
1	Outsider	Assault	Never Abort	Assault	Never Abort
3	Insider - One Try	No Try	-	Working Hrs. Diversion	Never Abort
4	Insider - Mult. Tries	No Try	-	Working Hrs. 10 kg.	Abort if Guards Detect
9	Insider - Minor Resources	No Try	-	Divert From Sampler	Never Abort
11	Manager - One Try	Disable Monitors	Abort if Guards Detect	Working Hrs.	Never Abort
12	Manager - Mult. Tries	No Try	-	Working Hrs. 10 kg.	Never Abort

Figure 1.7 Comparison of Adversary Strategy and Tactics for Two Designs

success are much lower with Design 4, so the adversary is better off not to risk being caught in the attempt.

Next, we consider whether or not to require a two-person rule for all plant operations. The model will produce several measures of performance by which this rule may be judged. These are listed in Figure 1.8 along with performance comparisons of designs with and without the two-person rule.

We assume the cost of the two-person rule to be \$500,000 per year. Compared to reduced consequences (valued at $\$15.4 - 5.8 = \$9.6 \cdot 10^6/\text{year}$), its benefits far exceed its cost. This conclusion holds, in the illustrative data set, for an order-of-magnitude range in tradeoff values between lives saved and dollar cost.

We next examine an integrated safeguards evaluation--substituting a real-time quantity estimation system for frequent physical inventories. Figure 1.9 shows results. For aggregated measures: C_D , C_S , and Total Cost (C_T), the design with quantity estimators is superior. The same holds true for system performance measures: frequency of attempt and expected SNM diverted. Notice, however, that the design with quantity estimators (Design 11) has lower probabilities of detection and identification than the design with frequent inventories (Design 14). This is because Adversary 4 is deterred by Design 11, but not deterred by Design 14. When Adversary 4 tries against Design 14 he or she has a good chance (.53) of being detected, which raises the overall probability of detection for Design 14. Notice, however, that a high probability of detection does not necessarily lead to the best overall system performance.

Our next sample decision involves setting a constraint on the expected quantity of SNM diverted each year. To examine the characteristics of this decision, we use the graph in Figure 1.10. The horizontal axis measures system performance for several different system designs, fourteen in this case. We measure overall performance in terms of the expected SNM diverted per year. Assume the regulator is setting a standard for this quantity. On the vertical axis, we measure three costs as a function of expected SNM diverted for fourteen designs: C_S , C_D and C_T .

	Design 4	Design 5
PERFORMANCE MEASURES	<u>Base Case</u>	<u>Required Two-Person Rule</u>
<u>Subsystem Performance</u>		
- Probability of Detection	.2	.5
- Probability of Identification	.2	.4
- Probability of Interruption	.2	.1
<u>System Performance</u>		
- Frequency of Attempt Per 1000 Years	17	3
- Expected Number of Tries/Attempts	6	1
- Expected SNM Diverted (g/year)	42	18
<u>Aggregated Performance Measures</u>		
- Diversion Cost (\$10 ⁶ /year)	15.4	5.8
- Safeguards Cost (\$10 ⁶ /year)	1.7	2.2
- Total Social Cost (\$10 ⁶ /year)	17.1	8.0

Figure 1.8 Evaluating the Two-Person Rule

	Design 14	Design 11
	Base Case Plus Frequent <u>Inventory</u>	Base Case Plus Quantity <u>Estimators</u>
PERFORMANCE MEASURES		
<u>Subsystem Performance</u>		
- Probability of Detection	.2	.1
- Probability of Identification	.3	.2
- Probability of Interruption	.2	.2
<u>System Performance</u>		
- Frequency of Attempt Per 1000 Years	17	14
- Expected Number of Tries/Attempts	6	7
- Expected SNM Diverted (g/year)	42	21
<u>Aggregated Performance Measures</u>		
- Diversion Cost (\$10 ⁶ /year)	15.4	5.9
- Safeguards Cost (\$10 ⁶ /year)	6.7	4.2
- Total Social Cost (\$10 ⁶ /year)	22.1	10.1

Figure 1.9 Comparison of Frequent Inventory and Quantity Estimation Monitors

Figure 1.10 Performance Graph: Cost Versus Expected SNM Diverted Per Year

The optimal level of performance is where total social cost (C_T) is lowest. This occurs at 6 g/year diverted. To the left of 6 g/year, diversion cost is lower, but the cost C_S of this decrease is substantial, causing C_T to rise rapidly.

Finally, we illustrate the results of a value/impact analysis for all 14 designs. This is shown in Figure 1.11, which is a simplified graphical representation of the relationship between diversion cost C_D (value) and safeguards cost C_S (impact). Each point on the graph corresponds to a different system design. Those that are on the boundary curve are most efficient: they attain the lowest diversion risk (highest value) at the lowest safeguards cost (impact). The curve illustrates that the risk of diversion can be lowered, but at increasingly high safeguards cost C_S . It tells, for example, an NRC commissioner or member of Congress how much it costs to reduce diversion risks to a particular level. Note that for the illustrative data set, it becomes very expensive to reduce diversion cost C_D below about $\$3 \cdot 10^6$ per year.

The only designs on the boundary in Figure 1.11 are 2, 3, 6, 7, and 8. The slope of the straight line in Figure 1.11 is -1, indicating 1 to 1 relative weighting between C_S and C_D . The point of tangency between the curve and the line gives the optional system -- Design 7 in this case. For the designs depicted, number 7 will be optimal over a wide range of weights: if the weight assigned to C_D by the decision-maker is anything more than 38% of the weight assigned to C_S , then Design 7 will be optimal. Notice that an extremely large weighting on C_D would be required before Design 3 would be optimal.

Similarly, this graph shows how the optimal design would change when numbers are varied in the Consequence Model. If the weighting between C_S and C_D is kept at 1 to 1, then C_D could be reduced by 62%, or increased almost limitlessly, and still Design 7 would be optimal. Of course, the robustness of Design 7 most likely would be less if there were more than fourteen designs plotted in Figure 1.11.

Figure 1.11 Value/Impact Tradeoff Curve: Safeguards Cost Versus Diversion Cost

The implication of this result is that the design decision is virtually insensitive to an increase in consequences. A doubling of the number of deaths and injuries resulting from a nuclear detonation in the Consequence Model would result in an 81% increase in the expected consequences of a 5 kg diversion; C_D for all designs increases by slightly less than 80%. This still does not make Design 3 nearly as attractive as Design 7.

Summary

The Aggregated Systems Model is a tool for setting performance criteria at a relatively high level. It integrates many factors in these criteria decisions, including: adversary information, facility safeguards response, diverted SMN consequences, and value/impact tradeoffs. However, it is meant to be an aid to, not replacement for, the regulators, who must integrate these factors and make the ultimate safeguards decisions.

A measure of deterrence has emerged from this analysis. It is rooted in the adversary's choice of how to attack a particular facility; if safeguards are strong enough, we hypothesize that the adversary will not make the attempt. A quantitative framework for determining this measure of deterrence is presented in this report.

Although an illustrative data set is presented, much of the information is judgmental. The ASM gives the regulator a tool to measure the sensitivity of the decision to this judgmental information. This insight can guide data collection efforts for the safeguards decision-making process.

II. AGGREGATED SYSTEMS MODEL DESCRIPTION

Introduction

Background

The Nuclear Regulatory Commission (NRC) is responsible for protecting the public against malevolent uses of special nuclear material (SNM) [1]. Lawrence Livermore Laboratory (LLL) is developing analytical procedures to assist NRC in meeting this objective. These procedures will assist in two areas:

- Setting safeguards criteria for facilities handling SNM
- Assessing individual plants to determine whether these criteria are satisfied.

This report describes the Aggregated Systems Model (ASM) developed by LLL and Applied Decision Analysis, Inc. (ADA) to aid the NRC in both areas of responsibility. Under the general title of setting safeguards criteria, we consider the following kinds of NRC decisions (arranged in order from limited to global scope):

- Component or procedure requirements: specifying the hardware and/or rules to be followed in a facility
- Integrated safeguards evaluation: making tradeoffs between safeguards components with different functions (e.g., material control versus physical security) to achieve the most effective safeguards system, possibly within cost constraints
- Performance-based regulations: choosing performance criteria for individual components or for overall safeguards systems
- Value/impact analysis: finding the socially optimal safeguards system with the best balance between the social need for safeguards and the social cost of the safeguards system.

The analytical models described here help assess individual plants in two ways:

- Preliminary assessments: making a first-cut assessment, using subjective data and the ASM, which will highlight crucial areas and hence guide resource allocations in detailed analyses
- Summarizing detailed assessments: aggregating the results to test sensitivity to changing parameters and to facilitate communication among regulators, operators, and designers.

Definition

The Aggregated Systems Model is a comprehensive quantitative tool for analyzing the types of safeguards decisions described in the opening paragraphs. It covers many diverse factors in these decisions, including adversary threats, safeguards responses, and the consequences of successful diversions. It represents, at an aggregate level, information that is usually developed by separate detailed analyses of these diverse factors. However, it contains only a small fraction of the data used by the detailed analyses. For a summary of other definitions used here, see the Glossary at the end of the report.

Relation to Other Work

An earlier ADA report [2] describes in detail a few uses of the Aggregated Systems Model in safeguards decision-making. The techniques are summarized in a report by LLL [3] and in a paper by Judd and Weissenberger [4]. These reports focus on aggregated systems modeling. Detailed assessment procedures are described in reports by LLL [5,6] and ADA [7]. The report by ADA [7] discusses the linkage between the Aggregated Systems Model and detailed assessments.

Report Organization

The remainder of Chapter 2 describes the analytical forms of the refined Aggregated Systems Model, including detailed descriptions of the adversary-facility interaction during an attempt. Chapter 3 describes data inputs for the illustrative analysis contained in Chapter 4. Chapter 5 gives conclusions and directions for further study.

Aggregated Systems Model Overview

Overview Figure

The major elements of the Aggregated Systems Model are shown in Figure 2.1. The Diversion Model contains data that characterize adversaries, their attempts, and the facility's safeguards system's response to those attempts. We frequently use the name "Diversion Model" to mean the combination of the Adversary and Facility models. The Adversary Model includes adversary characteristics and the adversary's choice of strategy and tactics. The Facility Model represents the safeguards system's ability to detect and interrupt the adversary's attempt.

The Consequence Model includes Public Consequence and Utility sub-models. The Public Consequence Model describes possible malevolent uses of stolen SNM and the consequences. The Utility Model represents the decision-makers' assignment of societal values to all possible outcomes and the explicit tradeoff between safeguards costs and the risks of diverted SNM.

While the Diversion and Consequence Models quantify diversion risk, the Safeguards Technology Model quantifies economic costs of safeguards. The meter at the right-hand side of Figure 2.1 shows the combined evaluation of safeguards costs and diversion risk.

Using the ASM

Recall the ASM uses listed at the beginning of this chapter:

- Setting safeguards criteria
 - Component or procedure requirements
 - Integrated safeguards evaluation
 - Performance-based regulation
 - Value/impact analysis
- Aiding detailed assessments
 - Preliminary assessment
 - Summarizing results.

We can conceptualize how the ASM will be used for each of these functions.

Figure 2.1 Aggregated Systems Model Overview

Component or procedure requirements and performance-based regulation of safeguards systems requires only a portion of the ASM: the Adversary, Facility, and Safeguards Technology models. The regulator can change the list of required components or procedures, and then the ASM will show how overall system performance changes. Or the regulator can change a performance-based regulation, and the ASM will show whether or not safeguards components in the system can meet the regulation. Examples of performance requirements are: probability of detection, probability of interruption, limitations on the expected amount of SNM diverted each year, etc.

Integrated safeguards decisions use the same three models. In this case, the ASM is used to compare and to choose combinations of different safeguards (such as physical security forces, material control systems, or accounting systems) that achieve a given level of performance. The model can show different ways to meet the performance requirement, some of which may be more advantageous to the operator than others. For instance, one could trade off improved real-time accounting equipment with frequent plant inventory shutdowns. Alternatively, one could observe the change in plant performance given new rules for upgrading safeguards.

Value/impact studies are designed to help decision-makers evaluate the social benefits and costs of changing safeguards requirements or system designs. The evaluation has three steps. The first step is analyzing how the designs or requirements affect the facility's performance against adversaries. In Figure 2.1, this analysis is accomplished in the box labeled Facility Model. Performance can be measured by the arrow labeled "SNM Diverted," which emanates from the Facility Model. The second step is to express the public consequences of the given performance level. For example, the ASM can describe how a tightening of security reduces the amount of SNM diverted and hence reduces the potential number of deaths or damage due to malevolent use of SNM. This translation occurs in the Consequence Model box shown in Figure 2.1. The third step involves balancing changes in public consequences (values) with the social costs (impacts) of the safeguards to achieve the changed consequences. The Utility Model can help in this tradeoff process. However,

if the decision-maker feels comfortable with implicit (rather than explicit) tradeoffs, the Utility Model need not be used for value/impact decisions. Notice that the value/impact analyses with the ASM use more of the models in Figure 2.1 than do other uses.

As an aid to detailed assessments, the ASM can use some or all of the models in Figure 2.1. In preliminary assessments, the model needs at least the Safeguards Technology and Facility Models, especially when evaluating performance against a specified adversary threat. The Adversary Model is used to consider the entire range of threats, rather than only one. The output of a preliminary assessment is a statement of the given facility's performance, measured as probabilities of detection, interruption, or expected SNM diverted.

When the ASM is used to summarize detailed assessments, more of the models may be used. If the goal is to show which facility components are the weakest links in safeguards, only the Facility and Adversary Models are needed. At the other extreme, a statement of the total social risk posed by the plant may require all of the models in Figure 2.1.

Differences from an Earlier Aggregated Systems Model

ADA's November 1978 report [2] presented both a general analytic methodology and an illustrative application of an early ASM to reactor fuel reprocessing facility safeguards against diversion of SNM. The model described here has some significant improvements. Therefore, before describing the model in more detail, we compare it in general terms with the earlier version.

Implicit in the earlier model were assumptions about various adversaries' modus operandi, for instance their diversion paths [8], target quantity of SNM, etc. However, sensitivity analysis using the earlier model showed the overall diversion risk was sensitive to repeated attempts by sophisticated insiders. These adversaries were assumed to make one or more attempts to divert large quantities of SNM. Ivan Sacks at LLL [9] showed that adversaries might have a higher probability of success diverting a large quantity of SNM in numerous small increments. Therefore, the Aggregated Systems Model was refined to analyze explicitly the adversary's choice of diversion strategy (modus operandi) and tactics.

The following adversary choices are now explicit in the ASM:

- Diversion Strategy
 - Diversion path
 - Number of tries
 - Quantity of SNM attempting to steal on each try
- Diversion Tactics
 - Abort the attempt
 - Abort a try if detected.

These are explained later in this chapter.

The adversary is assumed to make these strategic and tactical decisions based on maximization of the expected utility of the outcomes from the attempts. This "Adversary Utility Model" represents a significant improvement on the Adversary Model in the earlier ASM.

To allow the adversary to choose attempt frequency, a dynamic probabilistic model was needed to replace the earlier static version. The dynamic model represents another major refinement to the earlier version.

The new model also has a new event--adversary "identification"--in addition to the "detection" and "interruption" events in the earlier ASM. This primarily pertains to adversaries who make repeated tries to divert SNM. It is not enough for the facility to detect an attempt to divert SNM; the safeguards authority must identify the adversary to prevent further attempts. If there is no confirmation of an adversary's presence, the detection might be dismissed as a false alarm. Thus, before an adversary can be interrupted (captured) he or she must be both detected and identified.

Changes were also made to the Consequence Model described in the earlier report. The primary change is to condition consequences on adversary characteristics.

Assumptions in the Analysis

We conclude this overview with several assumptions that should be stated before discussing the model details. Some of those that pertain to this ASM as well as the earlier ASM are:

- The data are purely illustrative;
- Only diversion attempts are modeled; sabotage is not;
- The hypothetical facility is the test bed design [10];
- Consequences are limited to terrorist acts, and do not include international nuclear proliferation;
- The societal utility of lives, damage, and dollars is traded off in a linear fashion; the overall utility function for society is nonlinear; it is a step function dependent on amounts of SNM diverted.

Another important assumption is basing the adversary choice model on expected utility. An alternate approach is to assume adversary choices based on other criteria, such as minimizing the probability of detection, maximizing the probability of success, etc. Sensitivity of the results to some of these alternate approaches is discussed in Chapter 4.

An additional major assumption is the probabilistic nature of the repeated events model. We assume the probabilities of acquiring SNM, and of being detected, identified, or interrupted do not change from one attempt to the next, unless the adversary is identified. This allows us to construct a tractable Markov model to represent the dynamic and probabilistic nature of repeated attempts.

Finally, we assume the adversary preselects the diversion strategy and tactics before the first attempt, and this decision is not changed during the attempts. This understates the diversion risk somewhat, but we do not feel the understatement is significant. This "preselection" assumption greatly simplifies the dynamic programming problem posed by the repeated attempts.

Facility Model

The Facility Model represents the system designed to safeguard SNM. There are nine basic building blocks, called "components," in the current version of the ASM:

- Quantity estimators
- Process state monitors
- Personnel monitors
- Procedure monitors
- Stationary guards
- Roving guards
- Two-person rule
- Nominal accounting system
- Frequent physical inventories.

(Note that each of these building blocks is actually an aggregated system, composed of many hardware or software items. We leave the modeling of the hardware and software to detailed assessment procedures; the ASM deals with the performance of only highly aggregated systems.)

When referring to the safeguards system "design" we mean the list of components installed in the facility. Changing the design means adding, deleting, or substituting components.

The performance of the safeguards system is determined by the performance of individual components. Calculating system performance requires input data on component performance, and an algorithm for aggregating component performance to get safeguards system performance. The data and algorithm are discussed in Chapter 3.

One step in the algorithm is to aggregate component performance to an intermediate level called "subsystem" performance. Subsystems are collections of components that have similar performance characteristics. For instance, in the current version of the ASM we aggregate components into three subsystems:

- Material control
- Physical security
- Accounting.

The aggregation scheme is shown in Figure 2.2.

Figure 2.2 Elements of a Safeguards System Design

There are two reasons for the aggregation. These are explained in the next paragraphs. Our reasons are based on a need to condition system identification probabilities and adversary abort decisions on the detection event. In other words, we feel the probability of identification might be different if a guard detected the adversary as opposed to diversion detection by a process monitor. Similarly, the adversary might behave differently in the two cases. This motivates the conditioning. But because we also believe many detection events lead to similar abort decisions or identification probabilities, we condition on aggregated subsystems rather than on individual components. The requirements this places on the aggregation scheme is that all components in a subsystem must lead to the same identification probabilities or adversary abort/no abort decisions.

The first reason for aggregation relates to the safeguards authority's ability to identify the adversary and confirm the attempt when a detection has occurred. The probability of this event is conditioned only on "which subsystem detected the adversary," not on "which component detected." Note the requirement this places on the aggregated components: all components in a subsystem must lead to the same chances of identification. In other words, if an expert were assigning the probability that a given adversary, once detected, could be identified, the expert would need to know only whether or not one or more components in the subsystem made the detection; the expert would not need to know which component in the subsystem detected the adversary. If we relaxed this constraint in a model with nine components and three subsystems, we would have to condition the identification probabilities on the detecting component (with 512 combinations for each adversary) rather than on the detecting subsystem (with eight combinations for each adversary).

The second reason for aggregating components into subsystems is similar to the first. Recall that the current model allows the adversary to make strategic and tactical decisions. (These will be discussed in more detail in the next section.) One tactical decision includes aborting the try if detected by the safeguards. We condition this tactical decision on which subsystem detected the adversary, not on which component detected.

This places a second requirement on the aggregated components: detection by one or more components in a subsystem must lead to the same abort/no abort decision by the adversary. Relaxing this assumption again leads to 512 assessments instead of eight.

Since we assume identification probabilities depend on the adversary's abort/no abort decision, the number of identification assessments is actually double the numbers given earlier. Thus, there would be 1024 assessments for each adversary if conditioned on components and 16 if conditioned on subsystems.

Adversary Model

This section describes the types of adversaries currently in the model, the decisions they make about how to divert SNM, and the utility they assign to possible outcomes.

Adversary Types

Our earlier report [2] defined fourteen adversary types, whose characteristics are listed in Figure 2.3. Sensitivity analysis showed that (in the illustrative data set) six of these adversaries dominated the other eight in terms of their contribution to the expected quantity of SNM diverted. Therefore, only adversaries 1, 3, 4, 9, 11, and 12 were retained for this analysis. These are starred in Figure 2.3.

For all adversaries, we define an attempt as the existence of the adversary who wants to divert SNM. The attempt may consist of one or multiple tries, or, if the adversary is deterred, no tries at all.

Diversion Strategy and Tactics

As mentioned earlier, the adversary's modus operandi was implicit in the earlier model [2]. We now model explicitly the adversary's choice of diversion strategy (pathway to the SNM, number of tries, and the quantity to steal on each try) and tactics (abort the attempt or abort a try if detected). We assume the adversary makes these choices based on his or her own preferences for outcomes such as capture, partial success, complete success, or failure without capture (assumed to be the same as no attempt at all).

The diversion strategy characterizes the kind of attempt the adversary makes, whereas the tactic specifies decision rules for aborting. The most important element in the strategy is the diversion path, also called a monitor target set (MTS). These terms refer to the set of safeguards components that feasibly could detect the adversary on his or her route to the target SNM, while acquiring the SNM, or while leaving the facility with it. The diversion path or MTS will always be a subset of the components in the system design.

Number	Access	Resources			Desired Quantity	Tries	Attempt Frequency (Att./1000 Yrs.)	Percent of SNM Diverted
		Equipment	Authority	Collusion				
1*	Outsider	Major	No	No	Bomb	One	2	38
2	"	Minor	"	"	"	"	.02	.003
3*	Insider	Major	"	Yes	"	"	.29	4.7
4*	"	"	"	"	"	Multiple	2.6	33
5	"	"	"	"	Less	One	.03	.049
6	"	"	"	No	Bomb	"	.03	1.1
7	"	"	"	"	"	Multiple	.26	1.2
8	"	"	"	"	Less	One	.03	.006
9*	"	Minor	"	Yes	"	Multiple	13	13
10	"	"	"	No	"	"	1.4	.14
11*	"	Major	Yes	Yes	Bomb	One	.03	1.2
12*	"	"	"	"	"	Multiple	.15	7
13	"	"	"	"	Less	One	.07	.11
14	"	"	"	"	"	Multiple	.11	.18

*These adversary types are dominant

Figure 2.3 Adversary Characteristics

An important aspect of the adversary's strategic decision is whether or not to abort the entire attempt. With some diversion paths (strategies), there may be so many opportunities for detection that the adversary may be better off not trying. This may also be true for all strategies given a "secure" system design, so the adversary quits.

The adversary's tactical decisions are concerned with whether or not to abort the try if detected. We assume an insider's chances of identification are lower if he or she aborts. In covert attempts, this means they may be better off to abort. The tactics chosen by the adversary include a set of decision rules such as "abort if detected by Subsystem 2 but not if detected by Subsystems 1 and 3." Another might be "abort only if detected by both Subsystems 1 and 2." These tactics do not specify that every try will be aborted, but rather the conditions under which an abort will be in his or her best interest. If those conditions occur, then the try will be aborted.

We assume the adversary evaluates all possible strategies and tactics, and then chooses the combination that best serves his or her interest. These strategies and tactics are preselected; that is, the decisions are made before the first try, and we assume the adversary does not change strategy or tactics for the duration of the attempt.

Adversary Utility Model

The adversary's choice among strategies and tactics is made based on maximum expected utility. We assume three things matter to the adversary:

- Capture
- The status quo (no diversion)
- The quantity of SNM diverted.

These three things form the basis of the utility assessment. For an adversary who will divert the SNM in one lump sum, we need only be concerned with the utility of capture, $U(C)$, the utility of acquiring the desired amount, $U(S=D)$, and the utility of the status quo, $U(S=0)$. The status quo represents the case where no try is made or the case in which no SNM is acquired before identification occurs. An adversary who repeats

the attempt to steal many small quantities also has utility for partial success, $U(S=m)$, where m represents the number of small quantities obtained before the adversary is identified.

Further, we assume that for the multiple quantity case, the adversary's utility function for diverted SNM is linear, with the exception of the first and last increment. The last increment might have more value than any other increment if the whole quantity is necessary for some goal. Figure 2.4 shows the form of a typical utility function.

Assessment of the utility function is accomplished using the three lottery questions shown in Figure 2.5. Three numbers, P_1 , P_2 , and P_3 , are required for each adversary. Given these three numbers, the utility function is:

$$U(C) = 0.0$$

$$U(S) = \begin{cases} P_1 & S=0 \\ aS+b, & 1 \leq S \leq D-1; D>1 \\ 1.0 & S=D \end{cases}$$

$$\text{where: } a = \begin{cases} \frac{(1-P_1)(P_3-P_2)}{D-2}, & D>2 \\ 0, & D=2 \end{cases}$$

$$b = \begin{cases} P_2(1-P_1) + P_1 - a \end{cases}$$

The coefficients a and b are derived by the straight line between points $U(S=1)$ and $U(S=D-1)$:

<u>S</u>	<u>U(S)</u>
1	$P_2(1-P_1) + P_1$
D-1	$P_3(1-P_1) + P_1$

The assessment questions posed by Lotteries 1, 2, and 3 are easy to understand.

C = Capture state

C' = No Capture

S = Number of successful diversions

D = Desired number of successful diversions ($D > 1$)

Figure 2.4 Adversary Utility Function

	<u>Lottery</u>	<u>Outcome</u>	<u>Utility</u>
Lottery 1	Minimum Acceptable Probability of Success

Lottery 2	Utility of the First Increment
-----------	--------------------------------

Lottery 3	Utility of the Last Increment
-----------	-------------------------------

Figure 2.5 Assessment Lotteries for the Adversary Utility Model

For Lottery 1:

Given only two outcomes, total success and capture, what is the minimum probability of success needed to induce an adversary's attempt?

A high value of P_1 means a strong aversion to capture.

Lotteries 2 and 3 apply only to adversaries with repeated attempts.

For Lottery 2:

What is the minimum probability of success for which an adversary would take an "all or nothing" gamble rather than keeping only the first increment?

A high value of P_2 implies a high value of the first increment.

For Lottery 3:

What is the minimum probability of success for which an adversary would take an "all or nothing" gamble rather than keeping all but the last increment?

A high value of P_3 implies a low value of the last increment. Also, we require $P_3 \geq P_2$. If not, then

$$U(S>1) < U(S=1),$$

or the adversary is worse off with more SNM.

Capture need not be the worst outcome from the adversary's perspective, as it is in Figure 2.4. Assume for the moment that:

$$U(S=D) > U(C) > U(S<D),$$

where C, S and D are defined in Figure 2.4.

Rational behavior for this adversary dictates selecting a strategy with maximum probability $p(S=D)$ and minimum probability $p(S<D)$. It is likely that such a strategy would be an "all or nothing" attempt, with $D=1$. This adversary will choose a strategy with maximum probability of success. Since this decision criterion is examined as a sensitivity case in Chapter 4, we will assume that all adversaries have the general utility function shown in Figure 2.4. That is:

$$U(S=D) > U(0 \leq S<D) > U(C).$$

The linear utility function from $U(S=1)$ to $U(S=D-1)$ implies the adversary is risk neutral for quantities in this range. In other words, the adversary is willing to "play the averages" for incremental gains in the region. Depending on the adversary's risk preference, this may understate or overstate the utility of some strategies for the adversary. However, given the flexibility of utility assignments for $U(S=0)$ and $U(S=D)$, and the uncertainty as to the actual utility functions of adversaries, we believe the linear segment of the function is not a serious model limitation.

Adversary-Facility Interaction

The interaction between the adversary and the facility safeguards is complex; therefore, we have modeled the interaction in some detail. Two factors make the interaction complex. First, the adversary will optimize his or her strategy to find the weakness in the safeguards system. As new safeguards are installed to foreclose one diversion path, we assume the adversary will shift to the next weakest link. Second, an adversary may make repeated covert tries, perhaps waiting until the system is vulnerable due to safeguards component failures. A safeguards component that is "slow but sure to detect" may be valuable in preventing diversion by a repeating adversary.

This section describes first how we model the adversary's choice of strategy. We then discuss a dynamic probabilistic model of repeated adversary trials.

Schematic Decision Tree Description

Figure 2.6 is a schematic decision tree showing major decisions and outcomes in the adversary-facility interaction. Square nodes represent decisions: those with an "S" in the box are social decisions (perhaps made by the facility designer following NRC guidelines) and those with an "A" are adversary decisions. Circles represent probability nodes, where the decision-maker is uncertain as to which outcomes will occur. We show only two branches at each node, although there can be many.

The first decision we consider is the facility design--which safeguards components to choose. This decision is made without perfect foresight as to which adversaries (if any) will attack and whether the system can detect, identify, or interrupt the attempt.

Each adversary chooses strategy and tactics, as discussed earlier. This decision is the third node in Figure 2.6. These decisions consider the probabilities of detection, identification, and interruption leading to the outcome states at the right-hand side of the figure. For some adversaries, the detection/identification/interruption sequence of events is repeated for several tries.

Figure 2.6 Decision Tree: Adversary - Facility Interaction

As indicated in the Glossary, detection is defined as the transmission of an "attempt" signal to the safeguards authority. However, at that point the signal is undistinguished from a false alarm. "Identification" means the authority knows the alarm is real and who the adversary is, so that the guard force can be dispatched to confront the adversary. "Interruption" means no SNM crosses the plant boundary on a given try and the adversaries are captured. Because of this capture assumption, even if SNM were diverted on a previous try, no harm can come from it. If the adversary is identified but not interrupted, the attempt stops but the previously diverted SNM is still potentially lethal.

The identification event was implicit in ADA's earlier report [2]. However, when using a dynamic model with explicit repeated attempts, identification is a crucial event. In order for physical security to interrupt the sequence, timely identification of the adversary sequence must occur. If the attempt is detected, aborted, and the adversary not identified in time, then the attempt may be repeated. Non-identification is the key to repeated attempts, and therefore it is included as an explicit event.

We assume the adversary's perceived probabilities of detection, identification, and interruption are equal to those that would be assigned by the system designer. In other words, the designer and the adversary have the same state of information. This assumption could be relaxed easily.

We conclude this description of the schematic decision tree with a discussion of the sequence of events in a repeated attempt. These appear in Figure 2.6, and they are shown in greater detail but still in schematic form in Figure 2.7.

The structure in Figure 2.7 is repeated for each try. The nodes in the tree are all conditional on the adversary's chosen diversion strategy and tactics. We assume these are constant for all attempts.

We assume that an adversary, whose method of diversion is by repeated attempts, will continue trying until all desired SNM has been diverted or until he or she is identified. This is a result of the Adversary Utility

Outcome States

Adversary
Utility
(jth try)

Not repeatable

Success (S_u)
Failure (F_u)
Capture (C)

$$\begin{matrix} U(S_j) \\ U(S_{j-1}) \\ U(C) \end{matrix}$$

Repeatability (limited)

Success (S_{R1})
Failure (F_{R1})

$$\frac{U(S_j) + E[U(L)]}{U(S_{j-1}) + E[U(L)]}$$

Repeatability (unlimited)

Success (S_{R2})
Failure (F_{R2})

$$\frac{U(S_j) + E[U(U)]}{U(S_{j-1}) + E[U(U)]}$$

S = Cumulative increments of diverted SNM in j tries
(at least one try is successful)

$E[U(L)]$ = Expected utility of a limited number of future attempts

$E[U(U)]$ = Expected utility of an unlimited number of future attempts

Figure 2.7 Repeated Attempt Sequence Events (Schematic Tree)

Model, which assumes utility is assigned only to possessing SNM; no "disutility" is associated with having to try again. For this reason, the adversary decision at the left side of Figure 2.7 has only one branch: "Try Again."

Detection by the safeguards subsystems is assumed to be:

- Timely: with sufficient time to allow physical security to try to interrupt the sequences,
- Late: eventual detection, but after the attempt is concluded,
- Never.

Since component detection is aggregated to the subsystem level, when the adversary makes a decision on whether or not to abort, he or she knows which subsystems, if any, have detected the attempt.

The second adversary decision in Figure 2.7, the tactical abort/no abort decision, is also shown with only one branch. This is to highlight our assumption that the adversary preselects a decision rule for abort/no abort before the first attempt, and sticks with the rule for all attempts. The rule specifies an "abort" or "no abort" decision for each possible detection outcome. An example of part of one rule is:

Detection by Subsystem:

<u>1</u>	<u>2</u>	<u>3</u>	<u>Tactical Decision</u>
Timely	Timely	Not Timely	Abort
Timely	Not Timely	Not Timely	No Abort
Not Timely	Timely	Timely	Abort

If all three subsystems can give timely detection, then there are eight decision rules to be chosen before the first attempt is made. This set of eight decision rules is a tactic. Appendix A describes the tactical decision in more detail.

The next node in Figure 2.7 is the binary "acquire SNM" probability node. The adversary either acquires the desired increment of SNM or not. We assume the quantity of SNM sought on each attempt is the same. If the adversary aborts, no SNM is acquired.

Identification of the adversary by the safeguards authority can come timely, late or never. If interruption is to occur, both detection and identification must be timely.

Seven outcomes states are possible with every attempt. They are classified as "not repeatable," "repeatable for a limited time," and "repeatable" (unlimited). Within these three categories, there are at least two states representing success or failure to acquire SNM during the just-completed attempt. Figure 2.8 shows which repeated attempt sequence events determine the outcome state. The identification event determines repeatability, and the "acquire SNM" event determines attempt success or failure. The "Success" and "Failure" states in the not-repeatable category are both instances where the adversary is not captured, but because he or she is identified, the attempt cannot be repeated.

Finally, Figure 2.7 lists adversary utilities assigned to each state. As explained earlier in this chapter in Figure 2.4, the adversary utility function has two dimensions: capture and the quantity of diverted SNM. The adversary's utility at the completion of the j^{th} attempt is shown in Figure 2.7, depending on which state is obtained. For non-repeatable states, the utility is either $U(\text{capture})$ or $U(\text{cumulative SNM diverted})$. The utility of repeatable states is $U(\text{cumulative SNM diverted})$, plus the expected utility of future attempts. The maximum utility is $U(\text{desired quantity of SNM})$.

Extensive Form of the Decision Tree

Figures 2.6 and 2.7 showed the general form of events and decisions in the adversary-facility interaction. We now show explicitly the tree structure implied by the schematic decision tree. Using this extensive form of the tree, we will show how the dynamic probabilistic model is constructed.

To facilitate our description, we assume the adversary has pre-selected a diversion strategy. Also, we assume the decision rules for

<u>Outcome States</u>	<u>Attempt Sequence Events</u>
Not Repeatable	
Capture	Interruption
Failure	Not acquire SNM, timely identification, not interrupted
Success	Acquire SNM, timely identification, not interrupted
Repeatable (Limited)	
Success	Acquire SNM, late identification
Failure	Not acquire SNM, late identification
Repeatable (Unlimited)	
Success	Acquire SNM, never identified
Failure	Not acquire SNM, never identified

Figure 2.8 Relation Between Outcome States and Attempt Sequence Events

aborting the try (tactics) have been chosen. The spectrum of tactics ranges from "always abort," in which the try is aborted if any subsystem detects, to "never abort," in which the try continues regardless of which subsystem detects the try. For this example, we assume the adversary's tactic is "always abort" if detected.

Figure 2.9 shows the decision tree that the adversary faces on each try. The adversary begins the diversion and is or is not detected by each of the three subsystems. We assume that the adversary will know if he or she has been detected in a timely fashion: there will be an audible alarm, flashing lights, guards will appear, etc. If the adversary is detected by a subsystem or a particular combination of subsystems he or she will respond with the action dictated by the preselected tactic. In this case, he or she will abort the try when detected.

If the adversary aborts the try, obviously no SNM will be acquired. If the try is continued, the adversary will actually be able to acquire any SNM only part of the time. Depending on physical barriers, the adversary physically may be unable to acquire the SNM regardless of MC&A or physical security.

The resulting probabilities of identification depend on which of the subsystems detected the adversary and whether or not the try was aborted. Logically, if the adversary has no SNM in his or her possession, the probability that the safeguards authority will deduce that an attempt has been made will be lower. Perhaps the detection will be regarded as a false alarm. By always aborting if detected, an inside adversary who wishes to make numerous tries can minimize the probability of identification at the cost of not obtaining SNM on every try.

The next nodes represent identification and interruption. In order to interrupt the attempt and capture the adversary, the safeguards system must both detect and identify the adversary before he or she leaves the facility.

The outcome states corresponding to the tactic "always abort" are listed on the right in Figure 2.9. Notice that there are only a few branches that result in capture, and that the number of branches that yield any SNM is also small.

Figure 2.9 Decision Tree For Each Try

Figure 2.9 represents the decision tree for the tactic "always abort." For each strategy there are a variety of tactics that could be similarly depicted. Generating all possible tactics requires a lengthy explanation. This can be found in Appendix A.

A further complication, not shown in Figure 2.9, is the probability that a subsystem detects an adversary at a later time, after the completion of the try. If the adversary is identified after the late detection, he or she will be apprehended at the next SNM diversion try. Appendix A also explains how this factor is included in the calculations of ultimate success or capture probabilities for the adversary.

Markov Model

The tree in Figure 2.9 can be coalesced into a lottery over the outcome states shown in the right-hand column. Consider the simple case where the probability of late detection equals zero. Five outcome states result: capture (C), unrepeatable success or failure (S_U , F_U), repeatable success or failure (S_R , F_R). The probability of ending up in each of these states is the sum of the probabilities of each branch in Figure 2.9 that corresponds to that state. The first probability node in Figure 2.10 represents the coalesced lottery from Figure 2.9.

If the adversary is able to repeat the try, he or she faces exactly the same lottery again. The crucial assumption is that probabilities do not change with each repetition. Of course, if the adversary is captured or identified, the process stops. On the end of each repeatable branch, the same lottery could be attached. The second and third nodes in Figure 2.10 demonstrates how the decision tree could be repeatedly expanded. This quickly results in a large and intractable tree. Rather than attempting to solve this tree, we modeled the repeating adversary's iterative and probabilistic process as an equivalent Markov process [11]. Figure 2.11 presents this new model; note that the Markov states and probabilities correspond directly to the states and probabilities in the outcome lotteries in Figure 2.10. Also note that three states, F_U , S_U , and C are trapping states at which the process stops.

Figure 2.10 Repeating Lottery on the Outcome States From Each Try

Figure 2.11 Markov Model

Markov processes are useful for modeling not only because of their explicit consideration of both the probabilities and the dynamic aspects of the case, but also for their ease in solving for various results. Using the mathematical formulas associated with Markov processes, we can find the long-run probabilities of ending up in each of the outcome states. This calculation will produce a probability distribution over the ultimate outcomes of the repeated trial. For example, suppose an adversary were going to steal some number (D) of small increments of SNM (S), and the adversary would try until he or she had stolen D amounts, or were captured, or identified and unable to repeat. In any of these cases, the try would end. The Markov model would yield the probability of being in each of these stopping states, as well as the number of times the try was successful. The Markov model produces, therefore, a new lottery, which represents all ultimate outcomes of the repeated process in Figures 2.10 and 2.11. This is labeled "Outcome Lottery" in Figure 2.12. Compare Figures 2.12 and 2.6. Note that the "Outcome Lottery" in Figure 2.12 now replaces the three repeated nodes: detection, identification, and interruption in Figure 2.6.

Figure 2.13 contains the equations used to calculate the long-run probabilities for Figure 2.12. Section I on Figure 2.13 gives the probability of stealing D increments and the probability of being captured. The probability of being only partially successful-- $P(F)$ --is the sum of all the probabilities of stealing amounts less than D but being identified before D units were acquired. The distribution on the number of units of SNM diverted before identification occurs is found in Section III, corresponding to the probabilities of $S=D-1$, $S=D-2$, ... $S=1$, $S=0$.

Notice that in Figure 2.12 each outcome state has an adversary utility associated with it. The adversary can calculate his or her expected utility by multiplying the probability of each outcome times the utility of each outcome and summarizing all the products. This corresponds to the equation in Section II of Figure 2.13. For a given strategy, the adversary can compare the expected utility for this tactic with the utility of each of the other tactics, and choose the tactic with highest utility. This is the best he can do for that given strategy. Using this, he or she can then compare all the strategies,

Figure 2.12 Decision Tree Representing Final Outcomes From Repeated Trials

OUTCOMES AFTER INFINITE ITERATIONS

I. Long Run Probabilities:

$$P(D) = \left(\frac{P_1}{1-P_2} \right)^D \cdot \frac{(P_5(P_1+P_2) + (P_3+P_4))}{(P_1+P_2)(P_3+P_4+P_5)}$$

$$P(C) = \frac{P_5}{(P_3+P_4+P_5)} \left(1 - \left(\frac{P_1}{1-P_2} \right)^D \right)$$

$$P(F) = \frac{(P_3+P_4)}{(P_3+P_4+P_5)(P_1+P_2)} \left(\left(\frac{P_1}{1-P_2} \right)^D - \left(\frac{P_1}{1-P_2} \right)^{D-1} \right)$$

II. Expected Utility To Adversary:

$$EU = U(C) \cdot P(C) + U(D) \cdot P(D) + \sum_{m=0}^{D-1} P(m \text{ & } F) \cdot U(m)$$

III. Distribution on SNM Diverted:

$$P(m) = \frac{P_1^m \cdot (P_5(P_1+P_2) + (P_3+P_4))}{(1-P_2)^{m+1} (P_1+P_2)(P_3+P_4+P_5)}$$

IV. Expected Number of Tries:

$$E(n) = \frac{1}{P_3+P_4+P_5}$$

Number of Successful Tries:

$$= \frac{P_1}{P_3+P_4+P_5}$$

Number of Failures:

$$= \frac{P_2}{P_3+P_4+P_5}$$

Figure 2.13 Equations For Computing Markov Statistics

finding the strategy and associated tactic that will yield the maximum utility. Figure 2.12 shows the structure of the adversary's decisions.

Figure 2.13 contains one more interesting statistic. Section IV lists the equation giving the expected number of tries the adversary will make before he or she is identified by the safeguards system (and the process stops).

Although we did not discuss it here, we must consider the possibility of late detection and identification. Appendix B contains an explanation of how we included it in the formulation of the ASM. The equations used to solve for the long-run probabilities in the outcome lottery become much more complicated, but conceptually they merely build upon what was just explained.

Once it produces the optimal solution for the adversary, the model can be used to choose an optimal design, as shown in Figure 2.6. In addition, the information can be used by an NRC regulator determining how well a facility meets licensing or performance criteria. How well a system performs against an adversary who knows the best way to divert SNM is the most stringent test.

Another measure of performance is to assign social utility to each of the adversary's outcome states using the Consequence Model discussed in the next section. For any given system design, this model produces a probability distribution over consequences (such as deaths, damages, etc.). After assigning utilities to consequences, one can roll back the tree in Figure 2.6 to compute an expected utility for any particular design. Obviously, society will rate the utility of adversary capture very positively and will consider all the negative possible consequences associated with diverted SNM. Repeating the process for different designs identifies the design with the highest expected utility.

The Markov model also calculates the values of various performance measures that can be used for facility evaluation. Assuming that the adversary will use optimal strategy and tactics, system responses such as the probabilities of detection, identification, and capture, as well as the expected SNM diverted, can be calculated.

Consequence Model and Social Utility Model

Figure 2.14 shows a revised consequence probability tree with illustrative data. The following improvements have been made over the original Consequence Model in the earlier ADA report:

- Consequences are conditional on adversary;
- There is a probability that the SNM will be recovered before consequences occur;
- Weapon detonation consequences are a function of device yield;
- Hoax diversions followed by weapon threats are modeled;
- Injuries are postulated in addition to deaths;
- The "Intent" node has been expanded to include symbolic diversions (See Reference 12).

We have conditioned consequences on adversary characteristics to reflect the following dependences on adversary:

- Probability of recovery
- Probabilities over intents (weapon, extortion, sale, or symbolic diversion).

The probability of building a successful weapon may also be dependent on adversary, however given a successful 10 kg diversion, we assume the adversary is sufficiently competent. This assumption could be changed easily. Figure 2.15 shows the probabilities assigned to dependent events for the six adversary types. These assignments replace those in Figure 2.14 where appropriate. Notice that the probability of recovery equals 1.0 for Adversary 9. This adversary is merely out to prove that SNM can be stolen. He is, by definition stealing at most .3 kilograms, and therefore he will never have 5 kilograms, which is computationally equivalent to recovery with probability 1.0.

The consequences of weapon detonations are purely illustrative. However, a report soon will be issued by Dr. Dean Kaul at Science Applications, Inc. (SAI) [13] describing a model to predict the consequences of such events. Specifically, his model predicts immediate deaths and injuries due to blast, radiation, and thermal effects from a nuclear weapon detonation in two typical U.S. cities.

Quantity Diverted

<u>Adversary</u>	<u>5 Kg.</u>				<u>.01 - 5 Kg.</u>			
	<u>P</u> <u>(Recover)</u>	<u>P</u> <u>(Weapon)</u>	<u>P</u> <u>(Extort.)</u>	<u>P</u> <u>(Sale)</u>	<u>P</u> <u>(Recover)</u>	<u>P</u> <u>(Weapon)</u>	<u>P</u> <u>(Extort.)</u>	<u>P</u> <u>(Sale)</u>
1	.2	.75	.10	.10	.40	.75	.10	.10
3 & 4	.3	.50	.20	.20	.10	.50	.20	.20
9	1.0	-	-	-	.90	.10	.50	.10
11 & 12	.3	.10	.10	.50	.10	.10	.10	.50

Figure 2.15 Consequence Model Probabilities

SAI's model could be used to derive twenty of the specific consequence numbers in Figure 2.14. Using the model for sensitivity studies would allow conversion to probability distributions of the single point consequence estimates in Figure 2.14. Because SAI's results are forthcoming, we have not done any additional work on the consequences of these events. The assumption implicit in our probability assignments in Figure 2.14 is that these bomb consequences are the dominant effects in the Consequence Model.

Figure 2.16 shows the expected values of public and private consequences, as a function of quantity diverted. The public consequences are computed by rolling back the probability tree in Figure 2.14 with probabilities assigned in Figure 2.15. Plant consequences are derived from Figure 2.2 in ADA's earlier report [2]. The right-hand columns give the 3-point consequence utility function for each adversary and diversion quantity.

Adversary	Public Consequences			Plant Consequences			Total Consequences		
	Qty. Diverted (kg)			Qty. Diverted (kg)			Qty. Diverted (kg)		
	<u><.01</u>	<u>.01-5</u>	<u>>5</u>	<u><.01</u>	<u>.01-5</u>	<u>>5</u>	<u><.01</u>	<u>.01-5</u>	<u>>5</u>
1	$16P_e$	76	4696	13	13	20	$16P_e + 13$	90	4700
3 & 4	$16P_e$	84	4396	14	14	10	$16P_e + 14$	100	4400
9	$16P_e$	6	0	10	1	0	$16P_e + 10$	7	0
11	$16P_e$	22	890	10	10	0	$16P_e + 10$	32	890

$$P_e = 1 - P(\text{capture})$$

Figure 2.16 Expected Consequences by Adversary and Quantity Diverted (Equivalent $\$10^6$)

III. ILLUSTRATIVE DATA

Purpose

This example analysis is designed to illustrate application of the methodology to setting safeguards criteria for a hypothetical facility. To make the example realistic, the numbers were developed by a few members of the LLL-ADA project team. The data reflect the subjective judgment of those individuals; they were not developed by detailed analysis and they should not be regarded as accurate.

ASM Influence Diagram--Probabilistic Dependence

Important Dependence Assumptions

The ASM is a dynamic, probabilistic model, and as such has probability distributions for numerous variables in the model. Where two variables are dependent (for instance adversary type and the probability of detection), we made conditional probability assignments. We use the influence diagram in Figure 3.1 to show this dependence among variables.

Each square and circle in Figure 3.1 represents a decision and an uncertain event respectively. A single box denotes the regulator's or designer's decision; the double box represents the adversary's decision. Arrows between two circles or squares represent dependence in the model. Consider the identification event in the center of the diagram. The probabilities assigned to identification--"Timely, Late, or Never"--are conditional on the adversary, the adversary's tactic--abort or not--and whether detection is timely or not. However, the identification probabilities are independent of system design, the diversion strategy, and whether or not the adversary acquires SNM. Although in reality there may be some dependence on these other factors, we are making the assumption that all essential information for determining identification probabilities is contained in the detection, tactics, and adversary nodes.

Figure 3.1. Diversion Model Influence Diagram

Some important independence assumptions are:

- The adversary decides whether or not to abort, knowing that detection has occurred, but not knowing if he or she has been identified;
- Interruption probabilities depend only on identification, and not on how the adversary was detected;
- The acquisition of SNM depends on the adversary but not the diversion strategy; that is, all feasible strategies will allow the adversary to acquire the SNM with the same probability;
- The adversary knows the plant design, at least to the extent of which safeguards components are included;
- The adversary may choose not to try if the design is particularly secure;
- The system design influences the probability of detection, but not the probabilities of identification or interruption, in other words, changing the design can improve the system's ability to detect an adversary, but cannot change the probability of identification or interruption.

Detection: Aggregating Component Performance

Figure 3.2 shows additional detail for the detection event in Figure 3.1. The actions of an adversary are detected by safeguard components. In Figure 3.2, Subsystem 1, SS1, is composed of Components C11, C12, C13, and C14. The probability that any component will detect the adversary is independent of detection by any other component (in the same or different subsystem). This is indicated by the lack of arrows among components in the large component performance circle. The probability of detection by any subsystem is the probability that one or more of its components detect the adversary. Figure 3.2 also shows that subsystem detection probabilities are independent of detection by components belonging to other subsystems. Thus, subsystems can have no components in common. This probabilistic independence among subsystems sets a constraint on the analytical process of aggregating components.

Figure 3.2 also highlights the assumption that adversary tactics and identification are dependent on subsystem, not component, detection.

Figure 3.2 Influence Diagram with Detection System Detail

Making the independence assumption greatly simplifies the algorithm for computing the probability of detection by a subsystem:

$$P(SS_i \text{ detects}) = 1.0 - P(SS_i \text{ not detect})$$

$$P(SS_i \text{ does not detect}) = \prod_j (1.0 - P(\text{Component } ij \text{ detects}))$$

where i is the index for subsystems, and j is the index for components included in subsystem i .

Although it simplifies the calculation of subsystem detection probabilities, the independence assumption is not mandatory in the ASM. It could be relaxed in one of two ways:

- Direct specification of subsystem detection probabilities
- Specification of subsystem detection probabilities for all possible component configurations.

The first approach avoids altogether the use of components. Designs and adversary strategies are specified in terms of the subsystems included, rather than the components of subsystems.

The second approach requires computing subsystem detection probabilities external to the ASM. This would be done by an algorithm for handling the stochastic dependence among components. The external model would compute the probability of subsystem detection, where the subsystem contains only components that are both in the facility design and in the adversary's diversion path. This is done by considering every combination in every subset of components in the subsystem. For Subsystem 1 in the current model, which has 4 components, this requires 15 external calculations (assuming no late detection):

<u>Number of Calculations</u>	<u>Components in diversion path</u>
4	$C_{11}, C_{12}, \dots$ one at a time
6	$C_{11}, C_{12}, \dots$ two at a time
4	$C_{11}, C_{12}, \dots$ three at a time
1	$C_{11}, C_{12}, \dots$ four at a time
<hr/> 15	

If late detection were possible, 30 probabilities must be specified for Subsystem 1: P_D (Timely), and P_D (Never) for all combinations of components. As long as the number of components is relatively small, this latter approach is not formidable.

Safeguard System Designs

The influence diagram in Figure 3.2 shows a dependence between subsystem performance and the safeguards system design. As we have mentioned, system designs are combinations of components. For instance, three sample systems are:

<u>System Design</u>	<u>Components Included</u>
- Minimal design	C_{12}, C_{21}, C_{31}
- Maximal design	$C_{11}, C_{12}, C_{13}, C_{21}, C_{22},$ C_{31}, C_{32}
- Intermediate design	$C_{12}, C_{13}, C_{21}, C_{22}, C_{31}$

The system's design determines which components are included in the safeguards system, and, therefore, influences the subsystem probabilities of detection.

Figure 3.3 is a matrix showing the example safeguards system designs. The nonzero entries in the matrix designate components included in the design (one row of the matrix for each design). The numbers 1, 2, and 3 indicate subsystem number. Designs 1 through 4 comprise the base case and span the space of feasible designs. The remaining designs in Figure 3.3 are used for sensitivity analysis.

Assumed component costs are shown at the bottom of Figure 3.3. They are listed in thousands of dollars per year.

Adversaries

Adversary descriptions and illustrative attempt frequencies are shown in Figure 2.3. Figure 3.4 gives the probability that each adversary will acquire SNM.

Diversion Strategies

A diversion strategy includes a target quantity of SNM per attempt,

SUBSYSTEM:	<u>SS1 - Material Control</u>				<u>SS2 - Physical Security</u>			<u>SS3 - Accounting</u>			
COMPONENT:	<u>1.1</u>	<u>1.2</u>	<u>1.3</u>	<u>1.4</u>	<u>2.1</u>	<u>2.2</u>	<u>2.3</u>	<u>3.1</u>	<u>3.2</u>	<u>Inspections Per Year</u>	<u>Description</u>
DESIGNS:											
1	1	1	1	1	2	0	2	3	0	12	Testbed Design
2	0	0	0	0	2	0	0	3	0	6	Minimal
3	1	1	1	1	2	2	2	3	3	52	Maximal
4	0	1	1	0	2	0	0	3	0	6	Moderate
5	0	1	1	0	2	0	2	3	0	6	Physical Security (P.S.) - A
6	0	1	1	0	2	2	0	3	0	6	P.S. - B
7	0	1	1	0	2	2	2	3	0	6	P.S. - C
8	0	1	0	0	2	0	0	3	0	6	Material Control (M.C.) - A
9	1	1	0	0	2	0	0	3	0	6	M.C. - B
10	0	1	1	0	2	0	0	3	0	6	M.C. - C
11	1	1	1	0	2	0	0	3	0	6	M.C. - D
12	1	1	1	1	2	0	0	3	0	6	M.C. - E
13	0	1	1	0	2	0	0	0	0	6	Accounting - A
14	0	1	1	0	2	0	0	3	3	6	Accounting - B
Cost \$10 ³ Per Year:	2,450	600	600	2,000	300	400	500	300	5,000		

Figure 3.3 Safeguards System Designs

ADVERSARY

<u>Number</u>	<u>Description</u>	<u>P(Acquire SNM Attempt)</u>
1	Outsiders	.7
3	Insider; Major Equipment; One Attempt	.8
4	Insider; Major Equipment; Multiple Attempts	.8
9	Insider; Minor Equipment; Less than bomb quantity	.1
11	Insider; Major Personnel; One Attempt	.8
12	Insider; Major Personnel; Multiple Attempts	.8

P(Attempt) = 0.02

Figure 3.4 Adversary Probabilities for Acquiring SNM

an attempt frequency, and a diversion path or monitor target set. Recall that the diversion path is the collection of safeguards components that could detect the adversary during the attempt. If, for example, the target quantity is less than the detection threshold of the component, then the component is not in the diversion path. Figure 3.5 lists the diversion strategies for each adversary; a nonzero entry in any row means the component is included in the path. Only adversaries 4, 9, and 12 use multiple attempts. For these perpetrators, Figure 3.5 shows diversion strategies with varying frequency and quantity per attempt. Notice that when the quantity per attempt drops below 1 kg, (Strategy 4.1), the quantity estimation and process state monitors drop out of the diversion path. This reflects the assumption that their detection threshold is somewhat less than 1 kg. The "Do Not Try" strategy is not shown in Figure 3.5. However, it will be included in the model output in Chapter 4.

Detection

The influence diagram in Figure 3.2 indicates that detection probabilities for each component depend on the adversary. Figure 3.6 shows the other events considered in the component detection assessment. The first node on the form is an adversary decision whether or not to tamper with the component. For each component, we assume the adversary will evaluate both options, and then make the decision that maximizes his or her expected utility.

If the adversary tampers, and if there are tamper monitors on the component, then the first event considered is whether or not the tamper monitors are operational. If they are, then the adversary will be detected timely, late, or never. If they are not operational then they never detect the adversary.

Tampering takes on a special meaning in the case of Subsystem 2. There, the detectors are guards or other employees. We define tampering for Subsystem 2 as an attempt by the adversary to gain the collusion of noninvolved employees. An "operational" tamper system means the non-involved employee is aware that he or she is being approached. The tamper detection means the employee reports the contact.

SUBSYSTEM:	SS1 - Material Control				SS2 - Physical Security			SS3 - Accounting		Desired Qty(kg)	Increment (kg)	Freq. (Tries/Yr.)	Description
COMPONENT:	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2				
ADVERSARY STRATEGY													
1.1	0	0	0	0	2	2	0	0	0	10	10	1	Outsider Armed Attack
3.1	1	1	0	1	0	0	2	0	0	10	10	1	Normal Hrs. Diversion
3.2	1	1	1	1	2	2	0	0	0	10	10	1	Midnight Diversion
3.3	1	0	1	0	2	2	0	0	0	10	10	1	Breach Containment
3.4	0	0	0	0	2	2	0	0	0	10	10	1	Plant Mgr. Hostage
4.1	0	0	0	1	0	2	2	3	3	10	.3	52	Normal Hrs. Small Qty.
4.2	1	1	0	1	0	2	2	3	3	10	1	52	Normal Hrs. Large Qty.
4.3	1	1	0	1	2	2	2	3	3	10	10	52	Normal Hrs. Whole Thing
4.4	0	0	1	1	0	2	2	3	3	10	.3	52	Midnight Small Qty.
4.5	1	1	1	1	0	2	2	3	3	10	1	52	Midnight Large Qty.

Figure 3.5 Diversion Strategies

SUBSYSTEM:	SS1 - Material Control				SS2 - Physical Security			SS3 - Accounting		Desired Qty(kg)	Increment (kg)	Freq. (Tries/Yr.)	Description
COMPONENT:	1.1	1.2	1.3	1.4	2.1	2.2	2.3	3.1	3.2				
ADVERSARY STRATEGY													
9.1	0	0	0	1	0	0	2	3	3	.3	.3	52	During Wrk. Hrs. From Sampler
9.2	0	0	0	1	0	0	2	3	3	.3	.01	52	Small Amounts From Sampler
9.3	0	0	1	1	0	2	0	3	3	.3	.3	52	Midnight From Sampler
11.1	0	0	0	0	2	2	0	3	3	10	10	1	Disable MC&A
11.2	1	1	0	1	0	2	2	3	3	10	10	1	Normal Operation
11.3	1	1	1	1	2	2	0	3	3	10	10	1	Midnight
12.1	0	0	0	0	2	2	0	3	3	10	1	12	Disable MC&A
12.2	1	1	0	1	0	2	2	3	3	10	1	12	Normal Operation
12.3	0	0	0	0	2	2	0	3	3	10	10	12	One Shot Disable MC&A
12.4	1	1	0	1	0	2	2	3	3	10	10	12	One Shot Normal Oper.

Figure 3.5 (2) Diversion Strategies (Continued)

Figure 3.6 Probability Tree for Detection Assessment

The next node in Figure 3.6 is whether or not the component is operational. Given that it is operational, detection will be timely, late, or never. If both a tamper monitor and then the component itself detect late, the overall detection is late.

With the no tamper option, only the operation and detection by the component itself are considered. The braces on the right in Figure 3.6 show that the lotteries for both decisions can be collapsed to three-branch detection nodes: timely, late, and never.

We assume the adversary chooses the alternative with the lowest probability of timely detection. If the probability of timely detection is zero, as with records systems, then we minimize the probability of late detection. When there are repeated attempts, with nonzero probabilities of timely and late detection, we assume the adversary maximizes the probability of never being detected.

Figure 3.7 is the influence diagram used for detection probability assessment. Important assumptions shown in the diagram are:

- Obviously, component operation is influenced by tampering;
- The probability of component operation is not influenced by tamper detection, but only by the act of tampering itself;
- Component detection depends only on the component being operational, not on adversary tampering;
- The detection probability varies with the adversary, and whether or not NM is acquired. We assume SNM will be acquired by all adversaries in sufficient quantity to trigger detection. If the quantity is below thresholds, then the component would not be in the diversion path;
- The detection signal is treated the same whether it comes via the tamper monitor or the component itself.

The assessment of detection probabilities in Figure 3.6 requires seven numbers: p^{TMO} , p_T^{DT} , p_L^{DT} , p^{COT} , p^{CO} , p_T^{CD} , and p_L^{CD} for each component and each adversary. These are listed in Figures 3.8, along with comments on assessments.

Figure 3.7 Influence Diagram for Detection Probability Assessment

ADVERSARY/ COMPONENT	COMMENT	INPUT PROBABILITIES											
		Tamper Monitor P ^{TMO}	Tamper Detection		Component Operation		Component Detection		Computed Probs.				
			P ^{DT} T	P ^{DT} L	P ^{COT} T	P ^{CO} L	P ^{CD} T	P ^{CD} L	Tamper		No Tamper		Tamper?
									P(T)	P(L)	P(T)	P(L)	
1/All Components	Assume Detection						1.0	0			1.0	0	
3/1.1		.9	.7	0	.3	.9	.7	0	.71	0	.63	0	
3/1.2		.9	.7	0	.3	.9	.7	0	.71	0	.63	0	
3/1.3		.9	.7	0	.3	.9	.9	0	.78	0	.81	0	T
3/1.4		.9	.9	0	.1	.9	.95	0	.83	0	.86	0	T
3/2.1		.5	.2	.7	.2	.8	.5	.3	.2	.35	.4	.24	T
3/2.2		.8	.5	.1	.5	.9	.5	0	.55	.02	.45	0	
3/2.3		.8	.5	.1	.5	.9	.9	0	.67	.04	.81	0	T
3/3.1		.5	0	.5	.3	.7	0	.7	0	.42	0	.49	T
3/3.2		.9	0	.5	.5	.9	0	.95	0	.72	0	.86	T
4/1.1		.9	.7	0	.3	.9	.5	0	.69	0	.45	0	
4/1.2		.9	.7	0	.3	.9	.5	0	.69	0	.45	0	
4/1.3		.9	.7	0	.3	.9	.9	0	.78	0	.81	0	T
4/1.4		.9	.9	0	.1	.9	.95	0	.83	0	.86	0	T
4/2.1		.5	.2	.7	.3	.5	.3	.3		.37	.15	.15	
4/2.2		.8	.5	.1	.5	.9	.2	0	.46	.03	.18	0	
4/2.3		.8	.5	.1	.5	.9	.5	0	.55	.02	.45	0	

Figure 3.8. Detection Probability Assessments

ADVERSARY/ COMMENT COMPONENT		INPUT PROBABILITIES								Computed Probs.					
		Tamper Monitor P ^{TMO}	Tamper Detection		Component Operation		Component Detection		P ^{CD} P _T	P ^{CD} P _L	Tamper		No Tamper		Tamper?
			P ^{DT} P _T	P ^{DT} P _L	P ^{COT} P _{CO}	P ^{CO} P _{CO}	P ^{CD} P _T	P ^{CD} P _L			P(T)	P(L)	P(T)	P(L)	
4/3.1		.5	0	.5	.3	.7	0	.7	0	.42	0	.49	T		
4/3.3		.9	0	.5	.5	.9	0	.7	0	.66	0	.63			
9/1.1	Assume Not Skilled Enough To Tamper With SS1					.9	.8	0			.72	0			
9/1.1						.9	.8	0			.72	0			
9/1.3						.9	.9	0			.81	0			
9/1.4						.9	.95	0			.86	0			
9/2.1		.7	.5	.3	.5	.5	.5	.3	.44	.48	.25	.15			
9/2.2		.8	.5	.3	.5	.5	.2	0	.44	.42	.1	0			
9/2.3	P(T) Assigned Directly										.6	0			
9/3.1	Assigned.Compare To Adv.4										0	.7			
9/3.2	Same as 4										0	.63			
11/1.1	Assume No Tampering. If Adv.11 or 12 Tamper, then the component is <u>not</u> in Monitor Target Set, because Adv.11 & 12 have authority to disable Subsystem 1.					.9	.7	0			.63	0			
11/1.2						.9	.7	0			.63	0			
11/1.3						.9	.9	0			.81	0			
11/1.4						.9	.95	0			.86	0			

Figure 3.8 (2) Detection Probability Assessments

ADVERSARY/ COMPONENT	COMMENT	INPUT PROBABILITIES						Computed Probs.					
		Tamper Monitor P _{TMO}	Tamper Detection		Component Operation		Component Detection		Tamper		No Tamper		Tamper?
			P _{DT}	P _{DL}	P _{COT}	P _{CO}	P _{CD} P _T	P _{CD} P _L	P(T)	P(L)	P(T)	P(L)	
11/2.1		.5	.2	.7	.2	.5	.5	.3	.14	.82	.25	.15	T
11/2.2	Assume No Tamper					.9	.5	0			.45	0	
11/2.3	Assume Collusion	.8	.2	.2	.4	.7	.5	.3	.29	.33	.35	.21	T
11/3.1	Assigned Similar To Adv.3								0	.4	0	.5	T
11/3.2	" " " " "								0	.7	0	.8	T
12/1.1	Same as Adv. 11 Max. P(Never) Where Indicated:										.63	0	
12/1.2											.63	0	
12/1.3											.81	0	
12/1.4											.86	0	
12/2.1									.14	.82	.25	.15	
12/2.2	Max. P(N)										.45	0	
12/2.3									.29	.33	.35	.21	
12/3.1									0	.4	0	.5	T
12/3.2									0	.7	0	.8	T

Figure 3.8. (3) Detection Probability Assessments

Diversion Tactics

The model in its general form allows the adversary to make a tactical decision on whether or not to abort an attempt at three points, that is, after each of the three subsystem detects the attempt. However, detection for the adversary means timely detection, and currently only two subsystems (material control and physical security) can detect in a timely fashion.

This means there are only six possible tactics:

<u>Tactic</u>	<u>Description</u>
0	Do not try
1	Always abort if detected
2	Abort only if SS1 detects
3	Abort only if SS2 detects
4	Abort only if SS1 <u>and</u> SS2 detect
5	Never abort

See Appendix A for further discussion of tactics.

Identification

Figure 3.2 indicates that identification probabilities are dependent on the adversary, detection by subsystem, and on the adversary's abort tactics. Figure 3.9 shows the illustrative assignments for the probabilities of timely and late identification.

Interruption

The probability that the guards will interrupt the sequence depends on the adversary and timely identification. The illustrative probabilities are shown in Figure 3.10.

Adversary Utilities

Figure 3.11 is a copy of the form used for assigning adversary utility functions. Three input numbers are required: P_1 , P_2 , P_3 . These are tabulated for all adversaries in Figure 3.12. The forms of the curves are shown in Figure 3.13.

DETECTION BY SUBSYSTEMS			ABORT DECISION	ADVERSARY IDENTIFICATION PROBABILITY											
<u>1</u>	<u>2</u>	<u>3</u>		<u>1</u>		<u>3</u>		<u>4</u>		<u>9</u>		<u>11</u>		<u>12</u>	
				T	L	T	L	T	L	T	L	T	L	T	L
T	-	-	Y	1.0	0	.3	.3	.2	.2	.5	.3	.1	.1	.1	.1
T	T	-	Y	1.0	0	.5	.4	.5	.4	.6	.2	.2	.2	.2	.2
T	T	L	N	1.0	0	.8	.1	.8	.1	.8	.1	.5	.1	.5	.1
T	T	N	N	1.0	0	.8	.05	.8	.05	.8	.05	.5	.1	.5	.1
T	L	L	N	1.0	0	.6	.3	.5	.3	.5	.3	.3	.2	.3	.2
T	N	L	N	1.0	0	.6	.2	.5	.2	.5	.2	.3	.1	.3	.1
T	L	N	N	1.0	0	.6	.3	.5	.3	.5	.3	.3	.2	.3	.2
T	N	N	N	1.0	0	.6	.1	.5	.1	.5	.1	.3	.1	.3	.1
N	T	-	Y	0	0	.5	.3	.5	.3	.5	.3	.3	.2	.3	.2
N	T	L	N	0	0	.8	.1	.7	.1	.7	.1	.5	.1	.5	.1
N	T	N	N	0	0	.8	.05	.7	.05	.7	.05	.5	.05	.5	.05
N	L	L	N	0	0	0	.5	0	.3	0	.3	0	.2	0	.2
N	L	N	N	0	0	0	.3	0	.2	0	.2	0	.1	0	.1
N	N	N	N	0	0	0	0	0	0	0	0	0	0	0	0

T = Timely
 L = Late
 N = Never or No
 Y = Yes
 - = Not Applicable

Figure 3.9. Identification Probability Assessment

<u>Adversary</u>	<u>P (Interruption/Timely Identification)</u>
1	.20
3	.99
4	.99
9	.99
11	.99
12	.99

Figure 3.10 Interruption Probabilities

C = Capture State
S = Number of Successes
D = Number of Desired Successes

Figure 3.11 Adversary Utility Function Assessment Form

Adversary-Case	Name	P ₁	P ₂	P ₃	a	b
1.0	Base Case -- Adv. 1	.6	0	0	0	.6
3.0	Base Case -- Adv. 3	.7	0	0	0	.7
3.1	Low capture aversion	.2	0	0	0	.2
3.2	High capture aversion	.9	0	0	0	.9
4.0	Base case -- Adv. 4	.7	.2	.5	.09/(D-2)	.76-a
4.1	Low capture aversion	.2	.2	.5	.24/(D-2)	.36-a
4.2	High capture aversion	.9	.2	.5	.03/(D-2)	.92-a
4.3	Valuable first increment	.5	.2	.25	.03/(D-2)	.60-a
4.4	Valuable last increment	.5	.01	.05	.02/(D-2)	.51-a
9.0	Base case -- Adv. 9	.7	.3	.9	.18/(D-2)	.79-a
11.0	Base case -- Adv. 11	.9	0	0	0	.9
12.0	Base case -- Adv. 12	.9	.6	.9	.03/(D-2)	.96-a

Figure 3.12 Adversary Utility Function Parameters

SENSITIVITY ANALYSIS

- 1.0 Adv. 1
- 3.0 Adv. 3
- .-.-.- 3.1 Low Capture Aversion
- ..-.. 3.2 High Capture Aversion

- Base Case 4.0
- 4.1 Self-Risk Seeker
- .-.-.- 4.2 Self-Risk Averter
- ..-.. 4.3 Valuable 1st & Last Increments
- ...-... 4.4 Valuable Last Increment

Figure 3.13 Forms of Selected Adversary Utility Functions

IV. ILLUSTRATIVE ANALYSIS

Introduction

This chapter demonstrates types of analysis facilitated by the model using illustrative data. Because the data are illustrative, we do not draw substantive conclusions. However, we point out general insights to improve the user's understanding of the model and its sensitivity to a few key parameters. Together with data from an actual facility, these insights would give a decision maker a better understanding of the safeguards system's effectiveness against various types of adversaries.

We shall present four general types of model output:

- Adversary decision analysis
- Safeguards evaluation
- System performance graphs
- Sensitivity studies.

The first output evaluates the strategies and tactics available to each adversary. For a given system design, the model examines every alternative the adversary has and calculates the adversary's utility and other parameters such as detection probabilities.

The safeguards design evaluation produces an aggregated form of output. It repeats the adversary decision process for each system design, allows each adversary to choose his or her best alternative, and then produces system performance measures aggregated across all adversary types. This type of analysis focuses on the benefits of one design, or one performance standard, over another.

The final type of analysis, sensitivity analysis, demonstrates how the model results change when input data is changed. The example we give shows the results of changing adversary utility functions and systems designs. Numerous examples of other sensitivities are found in our earlier report [2].

Adversary Decision Analysis

Adversary Tactical Decisions

Figure 4.1a shows the most detailed level of model output for:

- System Design 1--testbed design (See Figure 3.3)
- Adversary 1--outsiders
- Diversion Strategy 1--armed attack (See Figure 3.5).

Figure 4.1a tabulates eight output parameters for all six tactics

Adversary 1 could choose. These parameters are:

- $E(U)$ Adversary expected utility
- P_{SD} Probability of successfully diverting the desired quantity of SNM (10 kg for Adversary 1)
- $E(N)$ Expected number of attempts by the adversary
- $E(SNM)$ Expected quantity of SNM diverted
- $E(S)$ The expected number of successes
- P_D Probability of detection on each attempt
- P_C Probability that the series of attempts will end in capture
- P_{ID} Probability that the adversary ever will be identified during the attempt(s).

Five tactics, ranging from "always abort if detected by any subsystem" to "never abort", are shown in Figure 4.2. In addition, tactic 0 represents the situation in which the adversary does not attempt the diversion.

These output measures indicate what is happening during each tactic and give insight into how a tactical decision is made. Look at the results in Figure 4.1a. The physical security force (Subsystem 2) will always detect Adversary 1 (See Figure 3.8). Tactics 1 and 3 are to abort if detected by the guards. Because guard detection is certain, Tactics (T) 1 and 3 will always mean failure: no SNM. Therefore, these have low adversary utility. With T2, T4, and T5, the adversary has .56 probability of success and .2 probability of capture. Detection by Subsystem 1 (SSI), the monitors, is superfluous, so T2, T4 and T5 are the same.

SYSTEM DESIGN 1;	ADVERSARY 1;		DIVERSION STRATEGY 1				ARMED ATTACK	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.480	0.000	1.000	0.000	0.000	1.000	.200	1.000
2	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3	.480	0.000	1.000	0.000	0.000	1.000	.200	1.000
4	.704	.560	1.000	.560	5.600	1.000	.200	1.000
5	.704	.560	1.000	.560	5.600	1.000	.200	1.000

Figure 4.1a Example Evaluation of Tactics -- Outsider

SYSTEM DESIGN 1;	ADVERSARY 4;		DIVERSION STRATEGY 3				ONE LUMP SUM	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.061	.054	3.821	.054	.544	.976	.936	.946
2	.116	.110	3.654	.110	1.096	.976	.882	.891
3	.286	.283	1.746	.283	2.826	.976	.713	.720
4	.293	.290	1.723	.290	2.902	.976	.706	.713
5	.312	.311	1.379	.311	3.113	.976	.687	.694

Figure 4.1b Example Evaluation of Tactics -- Insider

<u>Tactic Number</u>	<u>Abort if Detected by Subsystem:</u>
0	Don't try at all
1	1 or 2
2	1 only
3	2 only
4	1 and 2
5	Never

Figure 4.2 Adversary Tactics

The two lotteries represented by T1 and T5 are shown in Figure 4.3. The calculation of expected utility is also shown. The outcome states are those shown in Figure 2.12; adversary utilities are from Figure 3.12. The expected utilities for each tactic are shown in ovals, and they agree with those tabulated in Figure 4.1a. The best choice is to "never abort." To make the attempt but later abort is worse than not trying at all, because of the 0.2 probability of capture. The .56 probability of success comes from the .7 probability of acquiring SNM (Figure 3.4) times $(1-.2=.8)$ probability of no capture (Figure 3.10).

Figure 4.1a is a very simple case. Figure 4.1b, on the other hand, describes a more complex case. The situation is:

- Design 1---testbed
- Adversary 4---insider, multiple attempts possible
- Diversion Strategy 3---steal SNM in one 10 kg quantity.

Diverting one large quantity is somewhat simpler than an adversary stealing many small increments.

If the adversary makes a try, his or her utility shows T5--"never abort"--to be the best tactic because it has the highest value for expected SNM diverted and also the lowest probability of capture. T1 and T2 have significantly lower expected utility. This decreased utility depends on several factors. The expected amount of SNM diverted is smaller because so many of the tries are aborted. Abandoning the tries decreases the probability of capture on any given try; however, because the number of tries goes up, the probability of eventually being captured goes up too. Therefore, T5 is the best tactic. Nevertheless, the utility of T5 is less than the utility of not trying at all. The adversary will have the highest expected utility if he or she does nothing.

Adversary Strategic Decisions--General Description

Figure 4.4 evaluates all tactics for every diversion strategy (DS) considered by Adversary 3. Figure 3.5 showed these strategies:

- DS1 Normal hours diversion
- DS2 Midnight diversion

Figure 4.3 Adversary 1's Tactical Decision Problem

SYSTEM DESIGN 1;	ADVERSARY 3;		DIVERSION STRATEGY 1				NORMAL HOURS	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.493	.004	1.000	.004	.043	.992	.298	.301
2	.492	.013	1.000	.013	.130	.992	.302	.305
3	.372	.126	1.000	.126	1.259	.992	.523	.528
4	.370	.131	1.000	.131	1.306	.992	.528	.533
5	.261	.222	1.000	.222	2.223	.992	.722	.729

SYSTEM DESIGN 1;	ADVERSARY 3;		DIVERSION STRATEGY 2				MIDNIGHT	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.494	.002	1.000	.002	.023	.996	.296	.299
2	.493	.003	1.000	.003	.029	.996	.296	.299
3	.382	.274	1.000	.274	2.741	.996	.572	.578
4	.382	.274	1.000	.274	2.742	.996	.572	.578
5	.347	.295	1.000	.295	2.951	.996	.631	.638

SYSTEM DESIGN 1;	ADVERSARY 3;		DIVERSION STRATEGY 3				BREACH CONTAIN.	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.515	.037	1.000	.037	.375	.935	.281	.284
2	.514	.047	1.000	.047	.465	.935	.286	.289
3	.414	.297	1.000	.297	2.973	.935	.536	.541
4	.412	.300	1.000	.300	3.002	.935	.540	.546
5	.381	.324	1.000	.324	3.241	.935	.595	.601

SYSTEM DESIGN 1;	ADVERSARY 3;		DIVERSION STRATEGY 4				MGR. HOSTAGE	
TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	.804	.577	1.000	.577	5.766	.200	.099	.100
2	.791	.673	1.000	.673	6.733	.200	.153	.160
3	.804	.577	1.000	.577	5.766	.200	.099	.100
4	.791	.673	1.000	.673	6.733	.200	.158	.160
5	.791	.673	1.000	.673	6.733	.200	.158	.160

Figure 4.4 Adversary 3's Strategic Decision Evaluation

- DS3 Breach containment
- DS4 Take plant manager hostage.

The adversary only tries once. The system design is System 1--testbed design.

For DS1, Tactic 5 has the highest expected value for SNM diverted-- $E(SNM)$ --and a high probability of capture-- P_C . The high P_C relative to T4 outweighs the increase in $E(SNM)$ relative to T4, so T5 has lower expected utility than T4.

(The probability of detection is the same for all tactics under any given strategy. This is because tactics only influence responses after detection. The probabilities of ultimate identification and capture vary because some tactics cause more aborted attempts than others.)

Adversary 3 is nearly certain to be detected, but has rather low probabilities of being identified (See Figure 3.9), especially if he or she aborts (.3 if abort given SS1 detection, .6 if no abort). Notice the expected utility for DS1, given an attempt, is $E(U) = .493$, which is less than the utility of not trying: .7.

With strategy DS2--midnight diversion--the two-person rule is not in the diversion path, but the personnel monitors are (Figure 3.5). Notice with T3, T4, and T5, the $E(SNM)$ is 100 times greater than for T1 and T2. However, P_C is almost double; thus, the adversary's utility is lower for T5 than for T1 and T2. Because T1 and T2 have the same probability of capture, they yield about the same expected utility. The choice between T1 and T5 is shown in Figure 4.5. Once again, "not attempting" is the best strategy.

With Diversion Strategy 3--breach containment--we assume the personnel monitors and process state monitors drop out of the monitor target sets. In this case, the detection probability drops, $E(SNM)$ is slightly higher, and P_C is slightly lower, so $E(U)$ increases.

The fourth strategy--taking the plant manager hostage--shows that the adversary should make the attempt. All of the tactics have expected

Figure 4.5 Adversary 3's Choice of Tactics in Diversion Strategy 2.

utilities higher than that of not trying. Tactics 1 and 3, however, have the highest utility. Both of these involve aborting if the physical security force detects the try. Even though less SNM is expected to be diverted with T1 and T3, probability of capture decreases when the adversary chooses these two tactics. Thus the adversary should make the try but abort if he or she is detected by the guards. This last strategy has low P_{ID} even when the adversary doesn't abort. This is a result of the input data, which assume Adversary 3 is sophisticated and hard to identify, regardless of diversion strategy.

Figure 4.6 shows a less detailed form of results. Only the best tactic is given for each diversion strategy. This makes it easy to examine each adversary's choice of diversion strategy number. As we have discussed so far, for System Design 1 Adversary 1's best tactic is 5. The best tactic for each of Adversary 3's four strategies are given next. The "0" strategy corresponds to "no attempt."

SYSTEM DESIGN 1;		C(S)	6.6;	C(D)	5.8;	C(T)	12.4		
ADV. STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.1 *	5	.734	.560	1.000	.560	5.600	1.000	.200	1.000
3.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.1	1	.493	.004	1.000	.004	.043	.992	.298	.301
3.2	1	.494	.002	1.000	.002	.023	.996	.296	.299
3.3	1	.515	.037	1.000	.037	.375	.935	.281	.284
3.4 *	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0 *	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.1	5	.008	.000	1.714	.580	.174	.906	.990	1.000
4.2	5	.008	.000	1.612	.497	.497	.972	.990	1.000
4.3	5	.312	.311	1.379	.311	3.113	.976	.687	.694
4.4	5	.008	.000	1.600	.488	.146	.979	.990	1.000
4.5	5	.008	.000	1.580	.472	.472	.994	.990	1.000
9.0 *	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.1	5	.059	.053	1.522	.053	.016	.944	.938	.948
9.2	5	.007	.000	1.550	.056	.001	.944	.990	1.000
9.3	5	.100	.094	1.933	.094	.028	.973	.898	.907
11.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1 *	3	.929	.659	1.000	.659	6.594	.140	.042	.042
11.2	1	.812	.010	1.000	.010	.098	.986	.099	.100
11.3	1	.811	.001	1.000	.001	.011	.999	.099	.100
12.0 *	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12.1	5	.094	.086	5.296	3.445	3.445	.250	.905	.915
12.2	5	.010	.001	2.303	1.050	1.050	.988	.990	.999
12.3	5	.837	.837	2.036	.837	8.371	.250	.163	.164
12.4	5	.569	.569	1.701	.569	5.685	.988	.431	.435

Figure 4.6 Evaluation of All Adversaries' Strategies -- Design 1

Adversary 4's Strategies

Adversary 4 can make repeated tries at taking increments of differing sizes to accumulate 10 kg. The strategies are:

- DS0 Do not attempt to divert SNM
- DS1 Normal hours, .3 kg increment
- DS2 Normal hours, 1.0 kg increment
- DS3 Normal hours, 10.0 kg increment
- DS4 Midnight diversion, .3 kg increment
- DS5 Midnight diversion, 1.0 kg increment.

The best strategy shown in Figure 4.6 is DS0. For DS1 through DS5, the probability of detection, identification, and capture are all relatively high, so Adversary 4 has a difficult tradeoff between capture and getting the desired SNM. Though not shown in Figure 4.6, all tactics have about the same expected utility for strategies DS1, DS2, DS4, and DS5. However, DS3--taking the one large quantity--is preferred to the strategies taking small increments. Notice that for DS1 through DS5, the adversary should minimize the number of tries; minimizing P_D is not optimal. However, none of the strategies has a high enough utility to induce the adversary to try.

If a system can be designed so that an adversary is better off not making an attempt, the adversary effectively has been deterred. Adversary expected utility therefore serves as a measure of deterrence. If no attempt will be made, one might postulate that the system is safe enough to protect against Adversary 4.

Figure 4.7 shows the same type of results for System Design 2 that Figure 4.6 showed for System 1. Results for Adversary 4 show that the two diversion strategies that involve .3 kg increments have the same expected utility, as do the diversions involving 1.0 kg increments. The adversary's choice can be pictured as a decision tree, shown in Figure 4.8. For each strategy, the outcome probabilities and adversary preferences are assigned. The expected utility for each strategy is calculated, showing that the "one large quantity" strategy is best.

SYSTEM DESIGN 2; C(S) .5; C(D) 21.5; C(T) 22.0

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.1 *	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.1 *	5	.940	.800	1.000	.800	8.000	0.000	0.000	0.000
3.2	3	.804	.577	1.000	.577	5.766	.200	.099	.100
3.3	3	.804	.577	1.000	.577	5.766	.200	.099	.100
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.1	5	.248	.242	26.279	20.231	6.069	0.000	.750	.758
4.2	5	.828	.827	12.944	9.563	9.563	0.000	.172	.173
4.3 *	5	.894	.894	2.305	.894	8.942	.150	.105	.107
4.4	5	.248	.242	26.279	20.231	6.069	0.000	.750	.758
4.5	5	.828	.827	12.944	9.563	9.563	0.000	.172	.173
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.1 *	5	.816	.815	9.143	.815	.245	0.000	.183	.185
9.2	5	.008	.000	22.286	2.130	.021	0.000	.990	1.000
9.3	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
11.2 *	3	.980	.800	1.000	.800	8.000	0.000	0.000	0.000
11.3	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12.1	5	.101	.092	5.559	3.655	3.655	.250	.899	.908
12.2	5	.656	.653	11.320	8.264	8.264	0.000	.343	.347
12.3	5	.864	.863	2.429	.863	8.632	.250	.136	.138
12.4 *	5	.998	.998	2.238	.998	9.980	0.000	.002	.002

Figure 4.7 Evaluation of All Adversaries' Strategies -- Design 2

Figure 4.8 Sample Strategic Choice by Adversary 4 Given Design 2

We could construct a similar decision tree for each adversary's choice of diversion path as well as quantity per try.

Adversaries 9, 11, and 12 Strategies

Figure 4.6 shows that Adversary 9, with minor resources, has little chance of stealing even .3 kg. Again, the strategy with the minimum P_D is not optimal; the adversary should maximize utility. In Figure 4.7 under System Design 2, Adversary 9, like all adversaries, does make the attempt.

Adversaries 11 and 12 are plant managers. The diversion strategies for Adversary 11 are:

- DS1 Disable MC&A system (Subsystem 1)
- DS2 Divert as part of normal operations
- DS3 Midnight diversion.

Exploiting his or her authority over the MC&A system is the best strategy in Figure 4.6. However, the best tactic in that case is to abort if the guards catch on to what is happening. With DS1, the perpetrator faces almost no risk ($P_C = .042$) and the SNM is "there for the taking."

Adversary 12 has the option of repeated attempts, either for one large quantity or small increments:

- DS1 Disable MC&A--small increments
- DS2 During normal operations--small increments
- DS3 Disable MC&A--one large quantity
- DS4 Normal operations--one large quantity.

None of Adversary 12's strategies in Figure 4.6 yield a utility high enough to warrant an attempt, although disabling the MC&A System comes close to being worth an attempt.

Since Adversaries 11 and 12 are both plant managers, with 12 having more flexibility than 11, Adversary 12's best option should have higher utility than Adversary 11's. However, comparing results in Figure 4.6

shows this is not the case. This is because the model forces Adversary 12 to make repeated tries if the first is unsuccessful. The first try has high utility: $E(U) \approx .929$. Unfortunately, after the first try there is an increasing chance of late detection, identification, and capture. This chain of events, combined with the adversary's utility, which values SNM very little relative to not being captured, causes the utility to drop.

The expected number of tries may be quite large, especially when the probability of detection is low. Such cases are evident in Figure 4.7.

Safeguards Design Evaluation

Overview

Figure 4.9 summarizes optimal adversary strategies and tactics for each system design. We will discuss the merits of individual designs in the next paragraphs. However, some adversary-to-adversary comparisons are noteworthy. These are:

- Regardless of the system, Adversaries 1 and 11 will not be deterred; therefore, each system anticipates some risk of diversion;
- Given the frequency distribution over adversaries in Figure 2.3 the greatest contribution to overall E(SNM) is by Adversary 4. Thus, any system that reduces Adversary 4's E(SNM) will have generally lower C_D ;
- Adversary 11's E(SNM) changes drastically over designs 5, 6, and 7. This is because physical security varies, and the adversary is choosing different diversion strategies in response to the variations. System Designs 3, 6, and 7 have the best response to Adversary 11 because they all contain roving guards;
- Adversaries 4 and 9 have unlimited undetected tries in Design 13, which has no accounting system. They can divert all the SNM they desire, totally undetected. This demonstrates one benefit of a nominal accounting system.

Figure 4.10 shows the performance of each system design, without the adversary detail. The performance measures are:

- C_S Safeguards Cost ($\$10^6$ /year)
- C_D Diversion Cost ($\$10^6$ /year)
- C_T Total Cost ($C_S + C_D$)
- #ATT Frequency of an attempt per 1000 years
- #TRY Expected number of tries to steal SNM per attempt
- P_D Probability of ever detecting an adversary given an attempt

SYSTEM DESIGN 1; C(S) 6.6; C(D) 5.8; C(T) 12.4 TESTBED

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

SYSTEM DESIGN 2; C(S) .5; C(D) 21.5; C(T) 22.0 MINIMAL

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.1	5	.940	.800	1.000	.800	8.000	0.000	0.000	0.000
4.3	3	.894	.894	2.305	.894	8.942	.150	.105	.107
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.2	5	.980	.800	1.000	.800	8.000	0.000	0.000	0.000
12.4	5	.998	.998	2.238	.998	9.980	0.000	.002	.002

SYSTEM DESIGN 3; C(S) 12.0; C(D) 2.7; C(T) 14.7 MAXIMAL

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.3	1	.811	.001	1.000	.001	.006	.999	.099	.100
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

SYSTEM DESIGN 4; C(S) 1.7; C(D) 15.4; C(T) 17.1 (BASE CASE) MODERATE

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.3	2	.719	.718	2.622	.718	7.181	.532	.280	.283
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

SYSTEM DESIGN 5; C(S) 2.2; C(D) 5.8; C(T) 8.0 2-PERSON RULE PS-A

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

SYSTEM DESIGN 6; C(S) 2.1; C(D) 4.0; C(T) 6.1 ROVING GUARDS PS-B

ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.1	2	.641	.241	1.000	.241	2.406	.630	.187	.189
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.2	1	.814	.145	1.000	.145	1.446	.796	.112	.113
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 4.9 (1) Adversary Decisions for All System Designs

PS-C									
SYSTEM DESIGN 7; C(S) 2.6; C(D) 2.7; C(T) 5.3 2-P.RULE&ROV.G.									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.0	0	.600	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.3	1	.810	.011	1.000	.011	.113	.984	.101	.102
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC-A									
SYSTEM DESIGN 8; C(S) 1.1; C(D) 15.4; C(T) 16.5 NO PERS.MON.									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.3	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.3	3	.719	.718	2.622	.718	7.181	.532	.280	.283
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC-B									
SYSTEM DESIGN 9; C(S) 3.5; C(D) 5.9; C(T) 9.4 QTY. EST.									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC-C									
SYSTEM DESIGN 10; C(S) 1.7; C(D) 15.4; C(T) 17.1 SAME AS 4									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.3	3	.719	.718	2.622	.718	7.181	.532	.280	.283
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC-D									
SYSTEM DESIGN 11; C(S) 4.1; C(D) 5.9; C(T) 10.0 QTY. EST. & PERS. MON									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.1	5	.816	.815	9.143	.815	.245	0.000	.183	.185
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MC-E									
SYSTEM DESIGN 12; C(S) 6.1; C(D) 5.8; C(T) 11.9 ALL MC									
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9.0	0	.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 4.9 (2) Adversary Decisions for All System Designs

SYSTEM DESIGN 13;		C(S) 1.5;		C(D) 19.0;		C(T) 20.5		ACTG-A NO ACTG.	
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.1	5	1.000	1.000	41.250	33.000	9.900	0.000	0.000	0.000
9.1	5	1.000	1.000	13.000	1.000	.300	0.000	0.000	0.000
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

SYSTEM DESIGN 14;		C(S) 6.7;		C(D) 15.4;		C(T) 22.1		ACTG-B FREQ.INV.	
ADV.STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.704	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	3	.804	.577	1.000	.577	5.766	.200	.099	.100
4.3	5	.719	.718	2.622	.718	7.180	.532	.280	.283
9.1	5	.789	.788	8.866	.788	.236	0.000	.210	.213
11.1	3	.929	.659	1.000	.659	6.594	.140	.042	.042
12.0	0	.900	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 4.9 (3) Adversary Decisions for All System Designs

DESIGN	C(S)	C(D)	C(T)	#ATT	#TRY	P(D)	P(ID)	P(C)	CSNM	USNM
1	6.65	5.79	12.44	3.0	1.0	.45	.38	.11	5.99	18.0
2	.50	21.47	21.97	20.0	5.9	.10	.17	.13	4.05	81.1
3	12.05	2.67	14.72	2.0	1.0	1.00	.55	.15	2.80	5.6
4	1.70	15.40	17.10	17.0	6.6	.23	.24	.19	2.48	42.2
5	2.20	5.79	7.99	3.0	1.0	.45	.38	.11	5.99	18.0
6	2.10	3.98	6.08	14.0	7.4	.17	.24	.18	.87	12.1
7	2.60	2.68	5.28	2.0	1.0	.99	.55	.15	2.86	5.7
8	1.10	15.40	16.50	17.0	6.6	.23	.24	.19	2.48	42.2
9	3.55	5.90	9.45	14.0	7.4	.16	.23	.17	1.48	20.7
10	1.70	15.40	17.10	17.0	6.6	.23	.24	.19	2.48	42.2
11	4.15	5.90	10.05	14.0	7.4	.10	.23	.17	1.48	20.7
12	6.15	5.79	11.94	3.0	1.0	.45	.38	.11	5.99	18.0
13	1.50	18.99	20.49	17.0	13.9	.08	.07	.02	3.00	51.0
14	6.70	15.40	22.10	17.0	6.4	.23	.25	.21	2.48	42.1

Figure 4.10 Summary Performance Measures for All Designs

- P_{ID} Probability of ever identifying an adversary given an attempt
- P_C Probability of ever capturing an adversary given an attempt
- CSNM Expected quantity (kg) of SNM diverted per year given an attempt
- USNM Unconditional expected quantity (kg) of SNM diverted per 1000 years (or g/yr.)

Design Sensitivity

Fourteen designs are evaluated in Figure 4.10. Designs 1, 2, 3, and 4 span the range of system designs; the rest are sensitivity studies. These four are:

- 1 Testbed design (our approximation)
- 2 Minimal safeguards
- 3 Maximal safeguards
- 4 Moderate safeguards (less than testbed).

Of the four, the testbed design is best on a total cost basis. Adding the last few safeguards to get the maximal system (Design 3) costs \$5.4 million/year, but reduces diversion cost by only \$2.28 million/year.

While these four designs contain the lowest and highest diversion cost C_D and safeguards cost C_S , they do not contain the best overall system (based on C_T). This was found during the sensitivity analysis of Designs 5 through 14.

Design 7 is best overall, according to the C_T criterion. It depends more heavily on physical security than on MC&A systems, using only process monitors and personnel monitors, and it has a nominal accounting system to detect repeated attempts. It also costs 61% less than the testbed and its diversion costs are 54% lower. Notice, however, it does not rank first on either P_D or P_C minimization criteria.

Marginal Component Benefit

The information in Figures 4.9 and 4.10 can be used to evaluate the incremental benefit of individual safeguards components as well as safeguards subsystems. A sample comparison is shown in Figure 4.11.

COMPONENT EVALUATION

<u>SUBSYSTEM COMPONENT</u>	<u>MARGINAL IMPACT</u>		<u>MARGINAL VALUE</u>	
	<u>ΔC_s</u>		<u>ΔC_D</u>	
1. Material Control				
1.1 Quantity Estimation	+2.45		-9.50	
1.2 Process State	--	Required MC	--	
1.3 Personnel Monitors	+ .6		0	
1.4 Procedure Monitors	+2.0		-0.11	
2. Physical Security				
2.1 Stationary Guards	--	Required PS	--	
2.2 Roving Guards	+ .4		-3.11	
2.3 Two-Person Rule	+ .5		-1.30	
3. Accounting				
3.1 Nominal System	+ .2		-3.59	
3.2 Frequent Inventory	+5.0		0.0	

Figure 4.11 Marginal Evaluation of Safeguards Components

There are enough sensitivity cases in Figure 4.10 to evaluate most components. For instance, the first row in Figure 4.11 shows the net change in costs when the quantity estimation component (1.1) is added to a similar design without quantity estimation. (Design 8 is the base case; Design 9 has quantity estimation equipment.) Safeguards costs go up by $\$2.45 \cdot 10^6$ /year, and diversion cost drops by $\$9.5 \cdot 10^6$ /year. This is a net decrease in total cost, so for this base case the quantity estimation equipment is well worth the price.

Personnel monitors add nothing to facility security. According to Figure 4.10: detection doesn't matter for outsiders; Adversary 3 takes the plant manager hostage; Adversary 4 steals SNM during working hours when he or she is supposed to be near SNM; and Adversaries 11 and 12 disable all monitors. The only adversary for whom personnel monitors are in the diversion path is Adversary 9, and he or she gets less than .03 kg/year. Thus, personnel monitors do nothing to decrease total costs; therefore, they are not worth the investment for this illustrative data set.

In contrast, increasing physical security is beneficial. We assume stationary guards are required. Adding the two-person rule is marginally useful; adding roving guards is even more valuable.

The nominal records system has value because ultimately it stops repeating adversaries. Frequent inventories are assumed to be very expensive, and since much of the SNM is diverted by adversaries who only try once, they are not worth the improved information they provide.

Performance Graphs

Figure 4.12 plots the information contained in the first three columns and the P_D column in Figure 4.10 for selected system designs. The horizontal axis is P_D , and the vertical axis plots C_S , C_D , and C_T . Designs 3, 5, 6, 7, 8, 9, and 13 appear in Figure 4.12, although they are not all plotted on each cost curve. This is because the only designs plotted on each curve are those with minimum cost for that curve. The lower curve in Figure 4.12 is the minimum safeguards cost C_S for the set of designs included on the graph. The middle curve is diversion cost and the upper curve is total cost C_T .

Using the criterion of minimum total cost the best design is Design 7. This system has a probability of detection equal to .99. A probability of detection equal to 1.0 could be obtained by spending considerably more and using Design 3 rather than Design 7. However, this added detection performance has little value since diversion cost does not decrease much for Design 3.

Figure 4.13 graphs costs versus another performance measure--the expected amount of SNM diverted per year. The results here are similar to the previous case. Design 7 is the minimum point on the total cost curve. Design 3, with slightly less SNM diverted, costs $\$9.5 \cdot 10^6$ more. Note that the direction of the curve is reversed relative to Figure 4.12: costs rise greatly as more SNM is diverted.

We can also plot the alternate performance measures:

- P_{ID} Probability of identification
- P_C Probability of capture
- #ATT The expected number of attempts per 1000 years.

The probability plots will resemble Figure 4.12. The shape of the curve is reversed for #ATT, as in Figure 4.13.

Figure 4.14 is similar to Figure 5.5 in ADA's earlier report. On the horizontal axis is C_D , and C_S is on the vertical axis. All fourteen designs are plotted. This curve shows the tradeoff between safeguards cost and diversion cost. The only designs on the efficient frontier are 2, 3, 6, 7, and 8. As the relative weighting between C_S

Figure 4.12 Performance Graph: Cost

Cost Versus Expected SNM Diverted Per Year

Figure 4.14 Value/Impact Tradeoff Curve: Safeguards Cost Versus Diversion Cost

and C_D changes in computing C_T (it is 1 to 1 in this report), a different design becomes optimal. The slope of the straight line in Figure 4.14 is -1. The point of tangency between the curve and the line gives the optional system--Design 7 in this case.

Figure 4.14 also demonstrates the sensitivity of optimal design to C_D . If C_D is weighted more than 38% of C_S , then Design 7 will be optimal for the designs depicted. Notice that an extremely large weighting on C_D would be required before Design 3 would be optimal. This graph shows how the optimal design would change when numbers are varied in the Consequence Model. If the weighting between C_S and C_D is kept at 1 to 1, then C_D could be reduced by 62% or increased almost limitlessly, and still Design 7 would be optimal. Of course, the robustness of Design 7 most likely would be less if there were more than fourteen designs plotted in Figure 4.14.

The implication of this result is that the design decision is virtually insensitive to an increase in consequences. A doubling of the number of deaths and injuries resulting from a nuclear detonation in the Consequence Model would result in an 81% increase in the expected consequences of a 5 kg diversion; C_D for all designs increases by slightly less than 80%. This still does not make Design 3 nearly as attractive as Design 7.

Adversary Sensitivity

A sensitivity case that is likely to have impact on design decisions is to vary P_1 --the utility of "no try"--in the Adversary Utility Model. This, in effect, changes the adversary's attitude toward capture, and hence the choice of strategy and tactics. It also changes his or her willingness to make the attempt, which would influence the optimal level of plant safeguards if the criteria were based on adversary utility. Figures 4.15 and 16 show the results of a run with these assumptions:

- $P_1 = .1$ for all adversaries ("fearless" adversaries)
- P_2 and P_3 remain unchanged.

Figure 4.15 shows the strategy and tactics each adversary will use under the first four system designs. This figure can be compared with Figure 4.9, the circles indicating where decisions have been changed relative to Figure 4.9. With one exception, the circled numbers in the tactics column are all 5's. As the adversary values SNM more and capture less, he or she will abort fewer tries. Therefore the adversaries will be choosing Tactic 5, "never abort."

Figure 4.16 can be compared to Figure 4.10. These are the results for each system design if the adversary assigns the utility value of .1 to the status quo. Because more adversaries will attempt the diversion, the expected SNM diverted and the diversion cost have both risen. Even so, Design 7 remains optimal, showing that system design is relatively insensitive to adversary utility.

Summary

We have shown several illustrative results and insights produced by this type of analysis. These illustrations show how the model output changes because of varying input parameters. In addition, they demonstrate how various performance measures might be used in safeguards criteria-setting.

SYSTEM DESIGN 1; C(S) 6.6; C(D) 12.7; C(T) 19.3

ADV. STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.584	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	5	.690	.673	1.000	.673	6.733	.200	.158	.160
4.3	5	.311	.311	1.379	.311	3.113	.976	.687	.694
9.0	5	.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	5	.763	.745	1.000	.745	7.446	.140	.069	.070
12.3	5	.837	.837	2.036	.837	8.371	.250	.163	.164

SYSTEM DESIGN 2; C(S) .5; C(D) 21.5; C(T) 22.0

ADV. STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.584	.560	1.000	.560	5.600	1.000	.200	1.000
3.1	5	.820	.800	1.000	.800	8.000	0.000	0.000	0.000
4.3	5	.894	.894	2.305	.894	8.942	.150	.105	.107
9.1	5	.815	.815	9.143	.815	.245	0.000	.183	.185
11.2	5	.820	.800	1.000	.800	8.000	0.000	0.000	0.000
12.4	5	.998	.998	2.238	.998	9.980	0.000	.002	.002

SYSTEM DESIGN 3; C(S) 12.0; C(D) 10.6; C(T) 22.7

ADV. STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.584	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	5	.456	.445	1.000	.445	4.452	.560	.444	.448
4.3	5	.279	.279	1.339	.279	2.791	.980	.719	.727
9.0	5	.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11.1	5	.606	.591	1.000	.591	5.913	.527	.261	.264
12.3	5	.647	.646	1.798	.646	6.465	.587	.353	.356

SYSTEM DESIGN 4; C(S) 1.7; C(D) 18.2; C(T) 19.9

ADV. STRAT	TACTIC	E(U)	P(SD)	E(N)	E(S)	E(SNM)	P(D)	P(C)	P(ID)
1.1	5	.584	.560	1.000	.560	5.600	1.000	.200	1.000
3.4	5	.690	.673	1.000	.673	6.733	.200	.158	.160
4.3	2	.718	.718	2.622	.718	7.181	.532	.280	.283
9.1	5	.815	.815	9.143	.815	.245	0.000	.183	.185
11.1	5	.763	.745	1.000	.745	7.446	.140	.069	.070
12.3	5	.863	.863	2.429	.863	8.632	.250	.136	.138

○ Denotes changed decision from Figure 4.9

Case: $P_1 = .1$ for all adversaries -- change from Figure 3.12

Figure 4.15 Adversary Decisions -- "Fearless" Adversary Sensitivity

DESIGN	C(S)	C(D)	C(T)	#ATT	#TRY	P(D)	P(ID)	F(C)	CSNM	USNM
1	6.55	12.68	19.33	9.0	1.5	.63	.42	.33	6.03	54.2
2	.50	21.47	21.97	20.0	5.9	.10	.17	.13	4.05	81.1
3	12.05	10.62	22.67	9.0	1.4	.83	.55	.46	4.86	43.7
4	1.70	18.21	19.91	20.0	5.9	.27	.23	.18	3.50	69.9
5	2.20	15.55	17.75	20.0	2.1	.82	.64	.59	3.08	61.5
6	2.10	15.62	17.72	20.0	5.8	.36	.29	.25	2.89	57.9
7	2.60	13.39	15.99	20.0	1.9	.85	.70	.65	2.52	50.5
8	1.10	18.21	19.31	20.0	5.9	.27	.23	.18	3.50	69.9
9	3.55	15.89	19.44	20.0	5.8	.28	.25	.21	3.23	64.6
10	1.70	18.21	19.91	20.0	5.9	.27	.23	.18	3.50	69.9
11	4.15	15.89	20.04	20.0	5.8	.28	.25	.21	3.23	64.6
12	6.15	14.44	20.59	20.0	1.9	.63	.65	.61	2.95	59.0
13	1.50	21.81	23.31	20.0	12.2	.14	.08	.04	3.94	78.8
14	6.70	18.20	24.90	20.0	5.8	.27	.24	.20	3.48	69.6

Figure 4.16 Summary Performance Measures -- "Fearless" Adversary Sensitivity

V. CONCLUSIONS

This report is intended to document the dynamic version of the Aggregated Systems Model and to demonstrate its performance. ASM's primary advantages over earlier models are: (1) it represents more accurately adversary choices among strategies and tactics, and (2) it reflects the dynamic nature of repeated attempts. An assumption central to the model is that probabilities of detection, identification, and capture do not change over time; that is, neither the adversary nor the system gets smarter--with the exception of late detection or identification.

The analytic process has generated a potentially important useful measure of deterrence. The adversary is deterred when the expected utility derived from the diversion attempt is less than the utility of not trying. The deterrence measure depends upon adversary preferences and safeguards performance.

In this report, we have shown the use of various performance measures, in addition to expected adversary utility; these include P_D , P_{ID} , P_C and $E(SNM)$. The type of system being evaluated determines the choice among these measures: for example, an MC & A system evaluation might depend on P_D or $E(SNM)$; a physical security evaluation might depend on P_{ID} or P_C .

The next step in the development of the ASM should be to review the illustrative data set for the testbed design or to acquire data for a real facility. The analysis found in Chapter 4 should be repeated using more realistic data. This will provide a summary level analysis of a facility. The Aggregated Systems Model can then be used to summarize the results of detailed facility assessments and to set criteria for existing and future plants.

APPENDIX A

ADDITIONAL INFORMATION ON ADVERSARY TACTICS

For each diversion strategy, the number of tactics that an adversary must consider is a function of the number of subsystems that might give timely detection. Given n subsystems, there are 2^n possible combinations of subsystems that might detect the attempt. We call this combination of possible detections by subsystems a "detection condition." For each of the 2^n detection conditions, the adversary must decide whether or not to abort. Under one condition, the decision is obvious: clearly, an adversary would not abort if he or she were not detected. Thus, there are $2^n - 1 = y$ detection conditions for which an adversary must make an abort/no abort decision.

We define a tactic as a rule that says "abort" or "no abort" for each of these y possible conditions. There are 2^y different ways to say "abort" or "no abort" for the y conditions. Therefore, there are 2^y or $2^{(2^n - 1)}$ tactics for n subsystems. For three subsystems there are 128 tactics to evaluate.

A drawing, such as Figure A-1, is useful for identifying these possible tactics. This figure shows the conditions--events where a subsystem or combination of subsystems does or does not detect--and the adversary's decision on whether or not to abort under that condition. This tree illustrates a specific example with three subsystems. There are 2^3 , or eight, detection conditions, all but one of which is followed by an abort/no abort decision. This yields the 15 endpoints on the tree. (In general, there are $2^{n+1} - 1$ endpoints.)

A tactic can be specified by the set of endpoints that the adversary could reach by following the abort/no abort decisions for each condition. If, for example, he or she chooses the tactic "always abort if detected by any subsystem," he or she could end up at endpoint 1 (detection by SS1), endpoint 9 (detection by SS2), endpoint 13 (detection by SS3) or endpoint 15 (no detection). If detection by Subsystem 3 did not increase his or

her probability of identification and capture, the adversary would not abort if only SS3 detected, yielding the set of endpoints: 1, 9, 14, and 15. Similarly, endpoints 1, 10, 13, and 15 are possible if the adversary continues when only SS2 detects.

Obviously, the endpoints for a tactic do not include both the "abort" and "no abort" branches emanating from a particular detection condition. For example, an adversary cannot reach both endpoints 3 and 4, (which follow detection by all three subsystems) because one indicates aborting and the other continuing. Endpoint 15, on the other hand, is possible for every tactic. Of the other fourteen endpoints, only seven can be chosen for any one tactic. Every branch that says "abort" has a corresponding branch that says "no abort,"* and only one endpoint from each pair can be chosen. Therefore, there are $2^7 = 128$ possible tactics in this case.

Many of these 128 tactics are not logically consistent if detections by subsystems are independent. For example, a rational adversary would probably not choose to abort an attempt if detected by Subsystem 1 alone, but choose "no abort" if detected by both 1 and 2. Examination of the 128 tactics shows that all but 19 of them are logically inconsistent. All but these 19 can be eliminated from further consideration. A table of all logically consistent tactics is included in Figure A-2, which lists these nineteen possible tactics.

In the illustrative data set utilized in the model, only Subsystems 1 and 2 could detect in a timely fashion; Subsystem 3 (accounting) could only detect a diversion later, after a completed diversion. This reduced the number of feasible tactics to five. These are:

1. (1, 9, 15) Always abort if detected by any subsystem
2. (1, 12, 15) Abort if detected by Subsystem 1, but continue otherwise
3. (2, 8, 9, 15) Abort if detected by Subsystem 1, but continue otherwise

* Pairs are 1&8, 2&5, 3&4, 6&7, 9&12, 10&11, 13&14

<u>Tactic</u>	<u>Set of Possible Endpoints in Figure A-1</u>	
1.	1,9,13,15	- Always Abort
2.	1,9,14,15	
3.	1,10,12,13,15	
4.	1,10,12,14,15	
5.	1,11,12,14,15	
6.	2,6,8,9,13,15	
7.	2,6,8,9,14,15	
8.	2,6,8,10,12,13,15	
9.	2,6,8,10,12,14,15	
10.	2,6,8,11,12,14,15	
11.	2,7,8,9,14,15	
12.	2,7,8,10,12,14,15	
13.	2,7,8,11,12,14,15	
14.	3,5,6,8,10,12,13,15	
15.	3,5,6,8,10,12,14,15	
16.	3,5,6,8,11,12,14,15	
17.	3,5,7,8,10,12,14,15	
18.	3,5,7,8,11,12,14,15	
19.	4,5,7,8,11,12,14,15	- Never Abort

Figure A2 Possible Endpoints in Figure A1

4. (2, 8, 12, 15) Abort if detected by both Subsystems 1 and 2,
but continue otherwise
5. (5, 8, 12, 15) Never abort.

For each tactic, a decision tree similar to the one in Figure 2.9 could be constructed.

Late Detection

Notice that the tree in Figure 2.9 represents only the adversary's perceptions. In actuality, additional branches, not shown, could reflect the possibility of late detection; obviously, at the time of the diversion, the adversary cannot know whether he or she will be detected later. However, we assume the adversary does know (before he or she begins the attempt) the probabilities of timely and late detection, as well as the resulting probabilities of identification associated with each.

Since the adversary does not know whether or not late detection will occur, the best that he or she can do is to make the tactical decision after considering explicitly the possibility of late detection, identification, and interruption in addition to the possibility of never being detected. Figure A-3 shows a portion of Figure 2.9 representing the actual probability tree (for one subsystem) and the adversary's observable tree. It also shows the adversary's calculation of probabilities for the observable tree.

The Markov model discussed in Chapter 2 considers only the simple case in which no late identification occurs. In reality, late identification is extremely important: it allows the safeguards authority eventually to identify and stop the adversary from making repeated attempts. Therefore, late identification must be included in the Markov model. Figure A-4 shows what happens to the Markov process when late identification is added. We define "x" as the number of tries between the try on which the late identification process is initiated and the try on which the identification is made. The variable x is calculated as the frequency of inspections that might detect an adversary late, divided by the frequency of adversary tries.

PERCEIVED BY ADVERSARY

Figure A3 Actual and Perceived Probability Trees

Figure A4 Markov Model With Late Identification

The probabilities of being captured or identified remain the same as they were in the case with no late identification. Every repeatable try has some probability of never being detected, k_2 ; in this case, the adversary remains in the Markov process at the top of the page. But there is some probability, k_1 , that a repeatable try will lead to late identification. If an adversary does something to cause late identification, he or she will have x more tries before being identified and captured or identified but escaping capture--unless he or she stops because all desired material is acquired. In either case, after x more tries, the adversary drops into the process illustrated at the bottom of the page once the late identification process is triggered. The next $x-1$ tries appear exactly the same as a try in the regular Markov process. It is on the x^{th} try that probabilities change, sending the adversary to one of the three nonrepeatable states.

The calculation of long-run probabilities for the outcome lottery becomes a two-step process. For the first x tries, the process behaves as it did in the simple Markov model. But on the $x+1$ try and every following try, the probability of late identification (brought about by the try x time units before) must be considered. Each try before the $x+1$ try has a probability of $(p_1 + p_2)$ that it will be repeatable; that is, no identification has occurred. At each try after x , the probability of "not being identified" drops from $(p_1 + p_2)$ to $k_2(p_1 + p_2)$ and the probabilities of "capture" and "unrepeatable but not captured" states rise in proportion to k_1 , where k_1 is the portion of repeatable tries that will be identified late. Actual calculations of probabilities and performance measures follow in Appendix B.

APPENDIX B

MATHEMATICAL DERIVATION OF RESULTS

The derivation of these algorithms relies heavily on the use of Markov statistics and geometric transformations. Transforming a discrete function, $f(n)$, into a geometric summation, $f_g(z) = \sum_{n=0}^{\infty} f(n)z^n$, follows from the observation that the series expansion of $f_g(z)$ has unique coefficients $f(n)$. Because the expansion is unique, the relationship between the discrete function and the transform is also unique. For a more detailed discussion of geometric transforms (also known as z-transforms) consult texts such as Dynamic Probabilistic Systems by Ronald Howard [11].

First calculate the probability that exactly m units of SNM are stolen, given you are identified at try n .

$$p(s = m|n) = \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C + \binom{n}{m} P_S^m P_F^{n-m} P_I$$

The first term reflects the probability that m units were diverted on the first $n-1$ tries and capture occurred on the n th trial.

Note: Capture implies that no SNM was successfully diverted on that trial.

The second term reflects the probability that m units were diverted on n trials and identification but not capture occurred on the n th try. The adversary might or might not have diverted SNM on the n th trial; both are possible.

P_S = Probability of success on a try.

$P_F = 1 - P_S$ = Probability of failure on a try.

P_C = Probability of capture given identification.

P_I = Probability of escape given identification.

$P_R = 1 - P_C - P_I$ = Probability of no identification, allowing the attempt to be repeated.

The probability that n trials occurred, with identification occurring on the n th trial is:

$$P(n) = \begin{cases} P_R^{n-1} (1 - P_R) & \text{if } n \leq x \\ P_R^{n-1} k_2^{n-x-1} (1 - k_2 k_2) & \text{if } n \geq x + 1 \end{cases}$$

where x = number of trials until late identification could occur.

k_1 = proportion of repeatable attempts that will be identified.
late.

k_2 = proportion of repeatable attempts that will never be identified; $k_1 + k_2 = 1.0$

The probability that m units are diverted is:

$$\begin{aligned} P(s = m) &= \sum_{n=m}^{\infty} P(s=m|n) \cdot P_R^{n-1} (1 - P_R) \quad \text{Note that } \left(\frac{n}{m}\right) = 0 \text{ if } n < m \\ &+ \sum_{n=x+1}^{\infty} P(s=m|n) \cdot P_R^{n-1} k_2^{n-x-1} (1 - k_2 k_2) \\ &= \sum_{n=m+1}^x \left(\frac{n-1}{m}\right) P_S^m P_F^{n-m-1} P_C P_R^{n-1} (1 - P_R) + \sum_{n=m}^x \left(\frac{n}{m}\right) P_S^m P_F^{n-m} P_I P_R^{n-1} (1 - P_R) \\ &+ \sum_{n=x+1}^{\infty} \left(\frac{n-1}{m}\right) P_S^m P_F^{n-m-1} P_C P_R^{n-1} k_2^{n-x-1} (1 - k_2 P_R) + \sum_{n=x+1}^{\infty} \left(\frac{n}{m}\right) P_S^m P_F^{n-m} P_I \\ &P_R^{n-1} k_2^{n-x-1} (1 - k_2 P_R) \end{aligned}$$

Note that $\sum_{n=a}^{\infty} = \sum_{n=0}^{\infty} - \sum_{n=0}^{a-1}$

$$\sum_{n=x+1}^{\infty} = \sum_{n=m+1}^{\infty} \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right) - \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right)$$

$$\sum_{n=x+1}^{\infty} = \sum_{n=m}^{\infty} \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right) - \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right)$$

Using geometric transforms:

$$\sum_{n=m+1}^{\infty} \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right) = P_S^m P_R^m P_C k_2^{m-x}$$

$$\left(1-k_2 P_R\right) \sum_{r=0}^{\infty} \binom{r+m}{m} \left(P_F P_R k_2\right)^r, \text{ where } r=n-m-1$$

$$= \frac{P_S^m P_R^m P_C k_2^{m-x} \left(1-k_2 P_R\right)}{\left(1-P_F P_R k_2\right)^{m+1}}$$

$$\sum_{n=m}^{\infty} \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} k_2^{n-x-1} \left(1-k_2 P_R\right) = P_S^m P_R^{m-1} P_I k_2^{m-x-1}$$

$$\left(1-k_2 P_R\right) \sum_{j=0}^{\infty} \binom{j+m}{m} \left(P_F P_R k_2\right)^j, \text{ where } j=n-m$$

$$= \begin{cases} \frac{P_S^m P_I k_2^{m-x-1} P_R^{m-1} \left(1-k_2 P_R\right)}{\left(1-P_F P_R k_2\right)^{m+1}} & \text{if } m \neq 0 \\ \frac{P_I \left(1-k_2 P_R\right)}{\left(1-P_F P_R k_2\right)^{x+1} k_2^{x+1} P_R} - \frac{P_I \left(1-k_2 P_R\right)}{k_2^{x+1} P_R} = \frac{P_I P_F \left(1-k_2 P_R\right)}{\left(1-P_F P_R k_2\right)^x k_2^x} & \text{if } m = 0 \end{cases}$$

$$\begin{aligned}
\text{Then } P(S=m) &= \frac{P_S^m P_R^m P_C k_2^{m-x} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} + \frac{P_S^m P_I k_2^{m-x-1} P_R (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} \\
&- \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} k_2^{n-x-1} (1-P_R k_2) \\
&- \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} k_2^{n-x-1} (1-k_2 P_R) \\
&+ \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} (1-P_R) \\
&+ \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} (1-P_R) \\
P(S=m) &= \begin{cases} \frac{P_S^m P_R^{m-1} k_2^{m-x-1} (1-k_2 P_R) (P_C P_R k_2 + P_I)}{(1-P_F P_R k_2)^{m+1}} \\
+ \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} (1-P_R - k_2^{n-x-1} (1-k_2 P_R)) \\
+ \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} (1-P_R - k_2^{n-x-1} (1-k_2 P_R)) & \text{if } m \neq \emptyset \\
\frac{(1-k_2 P_R) (P_C + P_I P_F)}{(1-P_F P_R k_2)} + (P_F P_I + P_C) \sum_{n=1}^x P_F^{n-1} P_R^{n-1} (1-P_R - k_2^{n-x-1} (1-k_2 P_R)) & \text{if } m = \emptyset
\end{cases}
\end{aligned}$$

The probability that the adversary will divert all the SNM he or she wants is:

$P(m \geq S_D)$, where $S_D \times \text{Unit/Attempt}$ is the quantity he or she desires.

If the adversary diverts S_D units, he or she will be satisfied and discontinue the attempt.

$$\begin{aligned}
 P(m \geq S_D) &= \sum_{m=S_D}^{\infty} P(m) = \sum_{m=S_D}^{\infty} \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} \left(\text{where } T_1 = P_C P_R k_2 + P_I \right) \\
 &+ \sum_{m=S_D}^{\infty} \left(\sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} (1-P_R k_2^{n-x-1} (1-k_2 P_R)) \right) \\
 &+ \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} (1-P_R k_2^{n-x-1} (1-k_2 P_R))
 \end{aligned}$$

Using Transforms

$$\begin{aligned}
 \sum_{m=S_D}^{\infty} \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} &= \sum_{r=0}^{\infty} \frac{P_S^{S_D+r} T_1 P_R^{S_D+r-1} k_2^{S_D+r-x-1} (1-P_R k_2)}{(1-P_F P_R k_2)^{S_D+r+1}} \\
 &= \frac{P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{S_D+1}} \sum_{r=0}^{\infty} \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^r \quad \text{where } r = m - S_D \\
 &= \frac{T_1 P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} (1-k_2 P_R)}{(1-P_F P_R k_2)^{S_D+1}} \left(\frac{1}{1 - \frac{P_S P_R k_2}{1-P_F P_R k_2}} \right) \\
 &= \frac{T_1 P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} (1-P_R k_2)}{(1-P_F P_R k_2)^{S_D+1}} \left(\frac{1-P_F P_R k_2}{1-P_R k_2} \right) = \frac{P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} T_1}{(1-P_F P_R k_2)^{S_D}}
 \end{aligned}$$

Then, given that $S_D \neq 0$,

$$P(m \geq S_D) = \frac{P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} T_1}{(1-P_F P_R k_2)^{S_D}} + \sum_{m=S_D}^x \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} (1-P_R k_2^{n-x-1} (1-k_2 P_R))$$

$$+ \sum_{m=S_D}^x \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} \left(1 - P_R - k_2^{n-x-1} \left(1 - k_2 P_R \right) \right)$$

$$P(m \geq 0) = 1$$

Note: Because the limits on the inner sum are m and x where m is determined by the outer sum, the outer sum is bounded above by x also.

Next, the expected amount of SNM diverted is calculated.

$$\begin{aligned} E(m) &= \sum_{m=0}^{S_D-1} m P(S=m) + \sum_{m=S_D}^{\infty} S_D P(S=m) \\ &= \sum_{m=0}^{S_D-1} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 \left(1 - P_R k_2 \right)}{\left(1 - P_F P_R k_2 \right)^{m+1}} \\ &\quad + \sum_{m=0}^{S_D-1} m \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} \left(1 - P_R k_2^{n-x-1} \left(1 - k_2 P_R \right) \right) \\ &\quad + \sum_{m=0}^{S_D-1} m \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m-1} P_I P_R^{n-1} \left(1 - P_R k_2^{n-x-1} \left(1 - k_2 P_R \right) \right) \\ &\quad + \sum_{m=S_D}^{\infty} S_D \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 \left(1 - P_R k_2 \right)}{\left(1 - P_F P_R k_2 \right)^{m+1}} \\ &\quad + \sum_{m=S_D}^x S_D \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} \left(1 - P_R k_2^{n-x-1} \left(1 - k_2 P_R \right) \right) \\ &\quad + \sum_{m=S_D}^x S_D \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} \left(1 - P_R k_2^{n-x-1} \left(1 - k_2 P_R \right) \right) \end{aligned}$$

Evaluating the first single sum:

$$\begin{aligned}
 & \sum_{m=0}^{S_D-1} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} = \sum_{m=0}^{\infty} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} \\
 & \quad - \sum_{m=S_D}^{\infty} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} \\
 & \sum_{m=0}^{\infty} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} = \frac{T_1 (1-P_R k_2)}{(1-P_F P_R k_2) P_R k_2^{x+1}} \sum_{m=0}^{\infty} m \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^m \\
 & \quad = \frac{T_1 (1-P_R k_2)}{(1-P_F P_R k_2) P_R k_2^{x+1}} \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right) = \frac{T_1 P_S}{k_2^x (1-P_S k_2)} \\
 & \sum_{m=S_D}^{\infty} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} = \sum_{r=0}^{\infty} \left(r+S_D \right) \frac{P_S^{r+S_D} P_R^{r+S_D-1} k_2^{r+S_D-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{r+S_D+1}} \\
 & \quad \text{Where } m=r+S_D \\
 & = \sum_{r=0}^{\infty} r \frac{P_S^{S_D+r} P_R^{S_D-1+r} k_2^{S_D-x-1+r} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{r+S_D+1}} + \sum_{m=S_D}^{\infty} S_D \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} \\
 & \sum_{r=0}^{\infty} r \frac{P_S^{S_D+r} P_R^{S_D-1+r} k_2^{S_D-x-1+r} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{r+S_D+1}} = \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^{S_D} \\
 & \sum_{r=0}^{\infty} r \frac{P_S^r P_R^{r-1} k_2^{r-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{r+1}} = \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^{S_D} \frac{T_1 P_S}{k_2^x (1-P_S k_2)}
 \end{aligned}$$

Then

$$\sum_{m=0}^{S_D-1} m \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}} = \frac{T_1 P_S}{k_2^x (1-P_R k_2)} \left(1 - \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^{S_D} \right) - \sum_{m=S_D}^{\infty} S_D \frac{P_S^m P_R^{m-1} k_2^{m-x-1} T_1 (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}}$$

*Note that this last summation is the inverse of the second summation in the original equation and they will cancel.

Putting all the pieces together:

$$E(m) = \frac{T_1 P_S}{k_2^x (1-P_S k_2)} \left(1 - \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^{S_D} \right) + \sum_{m=0}^{S_D-1} m \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_C P_F^{n-m-1} P_R^{n-1} \left(1-P_C - \left(k_2^{n-x-1} (1-k_2 P_R) \right) \right) + \sum_{m=0}^{S_D-1} m \sum_{n=m}^x \binom{n}{m} P_S^m P_R^{n-1} P_I P_F^{n-m} \left(1-P_C k_2^{n-x-1} (1-k_2 P_R) \right) + \sum_{m=S_D}^x S_D \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_C P_R^{n-1} \left(1-P_C k_2^{n-x-1} (1-k_2 P_R) \right) + \sum_{m=S_D}^x S_D \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_I P_R^{n-1} \left(1-P_C k_2^{n-x-1} (1-k_2 P_R) \right)$$

Expected Utility

$$E(UTIL) = P(S \geq S_D) \cdot U(S_D) + U(C) \cdot \left(\sum_{m=0}^{S_D-1} P(S=m \text{ and } C) \right) \\ + \left(\sum_{m=1}^{S_D-1} U(m) \cdot P(S=m \text{ and } I) \right) + U(NONE) \cdot P(S=0 \text{ and } I)$$

Where $U(S_D) = 1$ = the utility of acquiring the desired SNM.

$U(m)$ = the utility of acquiring m units of SNM.

$U(NONE)$ = the utility of having no SNM but not being captured.

$U(C) = 0$ = the utility of being captured.

$P(\text{Capture})$ = probability of capture given identification.

I = adversary was identified but escaped.

$$\sum_{m=1}^{S_D-1} U(m) \cdot P(S=m \text{ and } I) = \sum_{m=1}^{S_D-1} a \cdot P(S=m \text{ and } I) + \sum_{m=1}^{S_D-1} b \cdot P(S=m \text{ and } I)$$

$$\sum_{m=1}^{S_D-1} a \cdot \frac{P_S^m P_R^{m-1} P_I k_2^{m-x-1} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} = \frac{a P_I P_S}{k_2^x (1-P_R k_2)} \left(1 - \left(\frac{P_S P_R k_2}{1-P_F P_R k_2} \right)^{S_D} \right) \\ - \sum_{m=S_D}^{\infty} a \cdot \frac{S_D P_S^m P_R^{m-1} k_2^{m-x-1} (1-P_R k_2)}{(1-P_F P_R k_2)^{m+1}}$$

(See derivation on Page 6)

$$\sum_{m=1}^{S_D-1} b \cdot \frac{P_S^m P_R^{m-1} P_I k_2^{m-x-1} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} = \sum_{m=0}^{\infty} b \cdot \frac{P_S^m P_R^{m-1} P_I k_2^{m-x-1} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} \\ - \sum_{m=S_D}^{\infty} b \cdot \frac{P_S^m P_R^{m-1} P_I k_2^{m-x-1} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} - \frac{b P_I (1-k_2 P_R)}{P_R k_2^{x+1} (1-P_F P_R k_2)}$$

$$= \frac{b P_I}{P_R k_2^{x+1}} \left(1 - \left(\frac{P_S P_R k_2}{1 - P_F P_R k_2} \right)^{S_D} \right) - \frac{b P_S (1 - k_2 P_R)}{P_R k_2^{x+1} (1 - P_F P_R k_2)}$$

Similarly

$$\sum_{m=S_D}^{\infty} \frac{a S_D P_S^m P_R^{m-1} k_2^{m-x-1} P_I (1 - P_R k_2)}{(1 - P_F P_R k_2)^{m+1}} = \frac{a S_D P_S^{S_D} P_R^{S_D-1} k_2^{S_D-x-1} P_I}{(1 - P_F P_R k_2)^{S_D}}$$

Then

$$\begin{aligned} E(UTIL) &= P(S \geq S_D) + \frac{a P_I P_S}{k_2^x (1 - P_R k_2)} \left(1 - \left(\frac{P_S P_R k_2}{1 - P_F P_R k_2} \right)^{S_D} - \frac{\left(\frac{P_S P_R k_2}{1 - P_F P_R k_2} \right)^{S_D-1} S_D (1 - P_R k_2)}{(1 - P_F P_R k_2)^{S_D}} \right) \\ &+ \frac{b P_I}{P_R k_2^{x+1}} \left(1 - \left(\frac{P_S P_R k_2}{1 - P_F P_R k_2} \right)^{S_D} \right) - \frac{b P_I (1 - k_2 P_R)}{P_R k_2^{x+1} (1 - P_F P_R k_2)} \\ &+ U(NONE) \cdot \frac{P_I P_F (1 - k_2 P_R)}{(1 - P_F P_R k_2) k_2^x} + U(NONE) \sum_{n=1}^x P_F^n P_R^{n-1} P_I \left(1 - P_R k_2^{n-x-1} (1 - k_2 P_R) \right) \\ &+ \sum_{m=1}^{S_D-1} a m + b \sum_{n=m}^x \binom{n}{m} P_S^m P_F^{n-m} P_R^{n-1} P_I \left(1 - P_R k_2^{n-x-1} (1 - P_R k_2) \right) \end{aligned}$$

The probability of eventually being identified is calculated exactly like the probability of success with the exception that no success occurs if the adversary is identified but escapes.

$$P(S_L = m) = \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-x-1} P_R^{n-1} (1-P_R) + \sum_{n=x+1}^{\infty} \binom{n-1}{m} P_S^m P_F^{n-m-1} P_R^{n-1} k_2^{n-x-1} (1-k_2 P_R)$$

$$= \frac{P_R^m P_S^m k_2^{m-x} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} + \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_R^{n-1} \left(1-P_R k_2^{n-x-1} (1-k_2 P_R) \right)$$

(See calculations on Pages B1 and B2)

$$P(m \geq S_D) = \sum_{n=S_D}^{\infty} P(S_L = m)$$

$$= \sum_{m=S_D}^{\infty} \frac{P_R^m P_S^m k_2^{m-x} (1-k_2 P_R)}{(1-P_F P_R k_2)^{m+1}} + \sum_{m=S_D}^{\infty} \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_R^{n-1} \left(1-P_R k_2^{n-x-1} (1-k_2 P_R) \right)$$

$$= \frac{P_S^{S_D} P_R^{S_D} k_2^{S_D-x}}{(1-P_F P_R k_2)^{S_D}} + \sum_{m=S_D}^x \sum_{n=m+1}^x \binom{n-1}{m} P_S^m P_F^{n-m-1} P_R^{n-1} \left(1-P_R k_2^{n-x-1} (1-k_2 P_R) \right)$$

(See calculations on Page B-5)

$P(\text{TRAP}) = 1 - P(m \geq S_D)$, i.e., if the adversary does not steal his or her desired amount and quit, the adversary will be identified - trapped - unable to repeat attempts.

GLOSSARY

This glossary is organized by the following major categories: Adversary Model, Facility Safeguards Model, Adversary-Facility Interaction, and Performance Measures. Within categories, the organization is usually chronological according to attempt sequence of events. The following list shows the major categories and the terms defined within each category:

ADVERSARY MODEL

Adversary
SNM
Attempt
Try
Strategy
Diversion path
Monitor target set
Tactic
Detection condition

FACILITY SAFEGUARDS MODEL

System design
Subsystem
Component
Definitions of all subsystems
and components in the
illustrative data set

ADVERSARY-FACILITY INTERACTION

Detection
Identification
Interruption
Timing
Acquire SNM
Divert SNM

ATTEMPT OUTCOMES

Capture
Success
Failure
Repeatable attempt
Not repeatable attempt

PERFORMANCE MEASURES

Safeguards cost
Diversion cost
Total cost
Expected annual attempts
Expected tries per attempt
Expected adversary tries
Unconditional expected
annual diversion
Conditional expected
annual diversion

PERFORMANCE MEASURES (CONTINUED)

Probability of detection given
attempt
Probability of identification
and capture given attempt
Expected adversary utility
Deterrence
Prevention
Response

ADVERSARY MODEL

Adversary: People who might attempt to divert (steal) special nuclear material (SNM). They are generally classified according to their group size, employee status, equipment resources, special knowledge or authority, and desired quantity of SNM. Earlier analyses referred to the adversaries in terms of "adversary action sequences" or "sequence categories."

SNM (Special Nuclear Material): Plutonium or highly enriched uranium from which an explosive or dispersal weapon could be made.

Attempt: The initiation of an adversary's plan to divert SNM. An attempt implies both the existence of an adversary and the initiation of his or her plan to divert. If the adversary exists but is deterred from initiating the plan, then no attempt has occurred. An attempt may be composed of several tries, and it may last as long as several weeks or months.

Try: An adversary's action to acquire and divert some quantity of SNM. Generally, a try lasts for less than one day.

Strategy (Number of Tries, Quantity per Try, Diversion Path): The elements of an adversary's plan: (1) the number of tries comprising the attempt, (2) the quantity to be acquired on each try, and (3) the diversion path. One alternative considered by each adversary is the "no attempt" strategy.

Diversion path: The set of safeguards the adversary encounters on his path to and from the SNM. Equivalent to MTS.

Monitor target set (MTS): The set of safeguards components the adversary will encounter during the try.

Tactic: A decision rule that governs the adversary's reaction to

detection by the safeguards system. Responses are: "abort
the try" and "no abort."

Detection condition: The set of safeguards subsystems that have
detected a particular adversary's try.

FACILITY SAFEGUARDS MODEL

System Design: A collection of safeguards components (aggregated into subsystems) installed in a facility.

Subsystem: A collection of similar safeguards components that have been grouped together for two reasons: 1) Adversary tactics (abort/no abort) depend on which safeguards subsystems (not individual components) detect him or her, and 2) the probabilities of adversary identification (by the safeguards authority) depend on which subsystem--not component--detects the adversary.

Component: The elemental unit of the safeguards system in the ASM. In reality, ASM components are composed of many monitors, guards, or accounting records. However, in the ASM, these safeguards elements are called "components" in the aggregate.

Subsystem 1 (SS1)--Material Control: The set of electronic safeguards components. SS1 detection usually results in an electronic signal sent to a control panel. This system does not identify the adversary without additional information. Detection can be timely, late, or never.

Component 1.1 (Quantity Estimators): Bubblers and other monitors that measure changes in liquid volume.

Component 1.2 (Process State): Instruments that measure pressure, temperature, and other physical characteristics of fluids in process. Also, sensors that give valve positions, fluid flow, etc.

Component 1.3 (Personnel Monitors): Floormats, infrared area monitors, and other instruments that detect the presence of individuals in an area.

Component 1.4 (Procedure Monitors): Computers that monitor personnel access, valve position changes, and other actions by operators.

Subsystem 2 (SS2)--Physical Security: Guards and other human safeguards elements. SS2 detection usually results from observing the adversary or observing that something is amiss physically. This system usually identifies the adversary. Detection can be timely, late, or never.

Component 2.1 (Guard Stations): Stationary guards, usually found at the plant gate or storage vaults.

Component 2.2 (Roving Guards): Self explanatory.

Component 2.3 (Two-Person Rule): Two operators present for all operations.

Subsystem 3 (SS3)--Accounting: Records and procedures to track the plant's SNM inventory. SS3 detection results from a discrepancy in the books. This system does not identify the adversary without additional information. Detection can be only late or never.

Component 3.1 (Nominal Accounting System): The accounting procedures required by NRC.

Component 3.2 (Frequent Inventory): Assay of all SNM in the plant. This extensive and expensive procedure requires plant shutdown for one to two months to flush all pipes and vats, shake down filter bags, etc.

Inspection Frequency: The minimum time interval between the adversary's try and the first possible late identification. This term applies to a complete system design, and not to individual subsystems or components. Inspections reveal discrepancies in records as well as physical security violations.

Safeguards Authority: The person, group, or computer with the authority to order intervention by the security force in order to interrupt the adversary's try. The safeguards authority is able to identify the adversary if given sufficient information.

ADVERSARY-FACILITY INTERACTION

Detection: The safeguards authority receives a signal (electronic or verbal) that one or more components have sensed a predefined diversion condition. The detection signal may be a false alarm.

Identification: The safeguards authority determines that an attempt is occurring or has occurred; i.e., detection was not a false alarm. In addition, the safeguards authority knows who the physical security force should confront in order to interrupt the diversion.

Interruption: The physical security force stops the diversion and prevents any more SNM from crossing the plant boundary. We assume this includes capturing the adversary and preventing malevolent use of any SNM diverted in the total attempt.

Timing for detection or identification (Timely, Late, Never):

- Timely -- Physical security has sufficient time to interrupt the try on which detection occurred.
- Late -- Detection or identification occurs after the try is complete but perhaps in time to interrupt a future try.
- Never -- The safeguards authority does not detect or identify the adversary on a particular try.

Acquire SNM: Being physically able to take possession of a quantity of SNM during a try. This does not necessarily mean leaving the plant with the SNM.

Divert SNM: The act of leaving the plant with a quantity of SNM.

ATTEMPT OUTCOMES

Capture: The adversary is apprehended and cannot use SNM for malevolent purposes.

Success: The adversary is not captured and diverts some SNM during his or her try.

Failure: The adversary is not captured, but fails to acquire SNM on a particular try.

Repeatable (Limited, Unlimited):

Unlimited: The adversary may try again indefinitely.

Limited: The adversary is identified late and has a limited number of tries before identification occurs.

Not Repeatable: The adversary is identified and will be captured if he or she shows up again at the plant.

PERFORMANCE MEASURES

Safeguards Cost (C_S): The annualized cost of safeguards systems.

This should be the incremental cost for adding safeguards components.

Diversion Cost (C_D): A single number comparable with C_S that

reflects the uncertain and potentially catastrophic consequences of the malevolent use of diverted SNM. It is expressed as an expected value for a one year period.

Total Cost (C_T): The sum of safeguards and diversion costs.

$$C_T = C_S + C_D$$

Expected Annual Attempts (#ATT): Expected number of adversaries who will initiate their attempt each year.

$$\#ATT = \sum_i F_i$$

where

F_i = frequency of attempts (not tries) by adversary i (e.g., 3 attempts/1000 years).

i = index of adversaries who are not deterred from making attempts.

#ATT is the expected number of adversaries who will initiate their plan in one year.

Expected Tries per Attempt ($E(T)$): This is the expected number of times adversaries will try to divert SNM during each attempt.

$$E(T) = \frac{\sum_i F_i E(N_i)}{\sum_i F_i}$$

Expected Adversary Tries ($E(N_i)$): Expected tries before Adversary

i has been captured or identified, or has diverted the desired quantity. This is conditioned on the attempt by Adversary i .

Unconditional Expected Annual Diversion (USNM): Expected quantity of SNM diverted from the facility each year.

$$USNM = \sum_i F_i E(SNM_i)$$

$E(SNM_i)$ = Quantity diverted by Adversary i over his or her complete attempt.

i = Again, only adversaries who are not deterred.

Conditional Expected Diversion (Given an Attempt) (CSNM): Expected quantity diverted given that an attempt (not try) has occurred.

$$CSNM = \frac{\sum_i F_i E(SNM_i)}{\sum_i F_i}$$

Probability of Detection Given Attempt (P_D):

$$P_D = \frac{\sum_i F_i P_D^i}{\sum_i F_i}$$

P_D^i = Probability that Adversary i is ever detected during complete attempt.

Probability of Identification and Capture Given Attempt (P_{ID} , P_C):

Defined analogously as P_D .

Expected Adversary Utility ($E(U^A)$): The expected value of the adversary's utility over attempt outcomes: capture, success, and failure. We assume the adversary will behave as if he or

she were maximizing expected utility. Thus, if the expected utility of the attempt is less than the expected utility of no attempt, the adversary will not make an attempt.

Deterrence: A suggested measure of deterrence is the expected number of attempts that are not made because the expected adversary utility $E(U^A)$ for not attempting is greater than the expected utility for attempting.

Prevention: A suggested measure is the Unconditional Expected Annual Diversion, (USMD).

Response: Suggested measures are the probability of detection (P_D) for material control and accounting subsystems and the probability of capture (P_C --Interruption) for physical security subsystems.

REFERENCES

1. C. A. Bennett, W. M. Murphy, T. S. Sherr, Societal Risk Approach to Safeguards Design and Evaluation, U.S. Energy Research and Development Administration, ERDA-7, June 1975.
2. Methodology and Preliminary Models for Analyzing Nuclear Safeguards Decisions, Applied Decision Analysis, Inc., November 1978.
3. B. R. Judd and S. Weissenberger, Decision-Making for Nuclear Safeguards, forthcoming.
4. B. R. Judd and S. Weissenberger, A Systematic Approach to Nuclear Safeguards Decision-Making, forthcoming.
5. M. Dittmore, F. Gilman, J. Lim, Facility X Vulnerability Analysis: Phase 1 Report, NRC report (classified), forthcoming.
6. I. Sacks, A. Parziale, M. Shrot, J. Long, A Structured Approach to the Assessment of Material Control and Accounting Systems, Lawrence Livermore Laboratory internal project memo, MC 78-203, March 1978.
7. T. R. Rice and S. L. Derby, Overview of the Structured Assessment Approach and Documentation of Algorithms to Compute the Probability of Adversary Detection, ADA, December 1978.
8. M. Maltese, K. Goodwin, J. Schleter, Diversion Path Analysis Handbook, National Bureau of Standards, October 1976.
9. I. Sacks, Adversary Success Probabilities with Multiple Attempts, LLL internal project memo, MC 78-473, June 22, 1978.
10. I. Sacks, et. al., Material Control System Design: Test Bed Nitrate Storage Areas, LLL, May 1978.

11. R. A. Howard, Dynamic Probabilistic Systems, Volumes 1 & 2, John Wiley & Sons, 1971.
12. C. W. Kirkwood, Modeling Potential Adversaries for Material Control and Accounting Systems, memo to S. Weissenberger, LLL, from Woodward-Clyde Consultants, July 6, 1978.
13. D. Kaul, et. al., Consequences of Adversary Actions in the Nuclear Power Fuel Cycle, Science Applications, Inc., forthcoming.

NOTICE

"Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore Laboratory under contract number W-7405-ENG-48."

"This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately-owned rights."

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.