# Physical Protection System Design Analysis against Insider Threat based on Game Theoretic Modeling

Kyo-Nam Kim [a], Young-A Suh [a], Erich Schneider [b], Man-Sung Yim [a]
[a]*Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, Republic of*
[b]*Nuclear and Radiological Engineering. Program, Department of Mechanical Engineering, the University of Texas at Austin, Texas, USA*
[*] *Corresponding author: charismak@kaist.ac.kr*

## 1. Introduction

Conventional tools assessing the security threats to nuclear facilities focus on a limited number of attack pathways defined by the modeler and are based on probabilistic calculations. They do not capture the adversary's intentions nor accounts for adversarial response and adaptation to defensive investments [1]. As an alternative way of performing physical protection analysis, use of a game theory has been suggested. This study explores the use of game-theoretic modeling of physical protection analysis by incorporating the implications of an insider threat. The defender-adversary interaction along with the inclusion of an insider is demonstrated using a simplified test case problem at an experimental fast reactor system. Non-detection probability and travel time are used as a baseline of physical protection parameters in this model. As one of the key features of the model is its ability to choose among security upgrades given the constraints of a budget, the study also performed cost benefit analysis for security upgrades options.

## 2. Game Theory

Game theory is an optimization method that models and manages risks from adversaries [2]. The adversary who wishes to reach a target and uses stealth to evade detection but cannot defeat a response force is assumed to have complete knowledge of the physical protection system of facility. The goal of the adversary is to make the most consequences by theft or sabotage at a facility.

The game theory model has the advantage of modeling an intelligent adversary without the user defining the adversary's actions. The interaction between defender and adversary is modeled as a two-person Stackelberg game. The optimal strategy of both players is found from the equilibrium of this game.

The expected consequence of an adversary's attack is a product of the consequence value of target and the probability of the successful attack.

## 3. Model Description

For a test case problem, we model a hypothetical fast reactor with a layered defense system [reference]. Example layout has two potential targets; Reactor Shutdown Cooling System (RSCS) and Fuel Cycle Facility (FCF). With this conceptual design of the facility, networks or directed graph of arcs and nodes are modeled with nodes representing locations and arcs representing paths of movement between two locations. Each arc is assigned a non-detection probability and travel time.

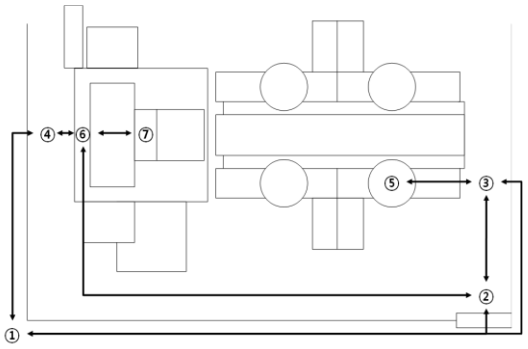


Fig 1. Network overlay of the facility

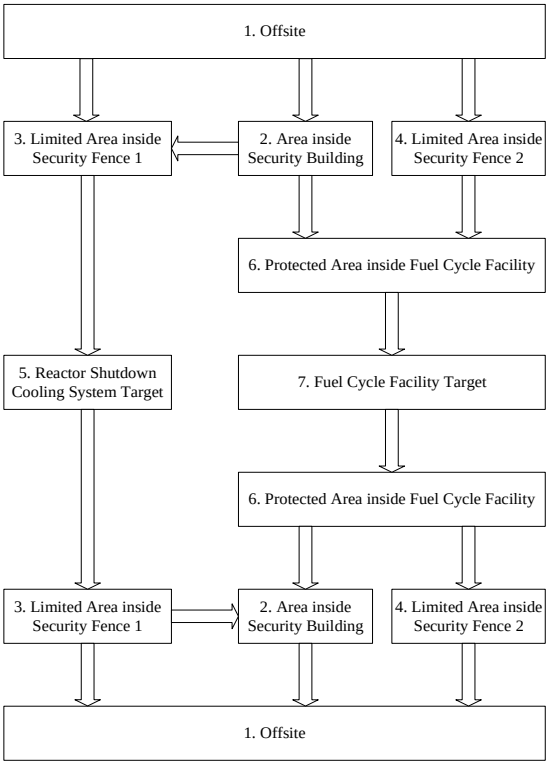


Fig 2. Mirrored network representation of facility

Note that the adversary succeeds not only by attacking the target but also by escaping the facility. We thus use a mirrored network in which the first half of the network contains pathways to the target and the second half (the

reflection) contains paths of egress. Travel times and non-detection probabilities can differ on these two halves of the network, and the ingress and egress networks themselves need not be perfectly symmetric.

The arc non-detection probabilities and travel times are populated by assessing a set of obstacles detecting or delaying an adversary traversing the arc and then estimating the probability of detection and time delay associated with each obstacle. It should be noted that the data in this model was constructed for demonstration purposes only.

Table I. Obstacles location and non-detection probabilities

| Path | Path Type | Fence | Personnel Portal | Stationed guards | Random Searches | Non Guard Personnel | Roaming Guards | Alarmed Detection | Video Surveillance |
|---|---|---|---|---|---|---|---|---|---|
| 1 → 2 | Personnel | | 0.99 | 0.95 | 0.97 | | | | 0.97 |
| 2 → 1 | Personnel | | | | | | | | 0.97 |
| 1 ↔ 3 | Fence | 0.91 | | | 0.97 | | | 0.91 | 0.91 |
| 1 ↔ 4 | Fence | 0.91 | | | 0.97 | | | 0.91 | 0.91 |
| 2 ↔ 3 | Limited Area | | | | 0.95 | | 0.97 | | |
| 2 ↔ 6 | Limited/ Protected Area | | | | 0.91 | | 0.96 | 0.91 | 0.80 |
| 3 ↔ 5 | Attack RSCS | | | | 0.77 | 0.97 | 0.84 | 0.80 | 0.74 |
| 4 ↔ 6 | Protected Area | | | | 0.88 | | 0.95 | 0.84 | 0.80 |
| 6 ↔ 7 | Attack FCF | | | | 0.80 | 0.97 | 0.88 | 0.84 | 0.74 |

Table II. Non-detection probabilities and travel times data

| Path | Non-Detection Probability | Travel Time |
|---|---|---|
| 1 → 2 | 0.885 | 40 |
| 2 → 1 | 0.970 | 10 |
| 1 ↔ 3 | 0.731 | 20 |
| 1 ↔ 4 | 0.731 | 20 |
| 2 ↔ 3 | 0.922 | 40 |
| 2 ↔ 6 | 0.636 | 80 |
| 3 ↔ 5 | 0.371 | 60 |
| 4 ↔ 6 | 0.562 | 40 |
| 6 ↔ 7 | 0.424 | 60 |

### *3.1. Model assumptions*

This study builds on a previous study by [4]. While the previous research work contains 6 model assumptions in greater detail, it does not consider the insider threat. We thus revise the model with the consideration of an insider while making additional assumptions as follows.

A. Type: individual with authorized access to a facility and system
B. Capabilities:
   i. Knowledge – layouts / security measures, vulnerabilities
   ii. Skills – ability to neutralize security measure
   iii. Number – 1 insider
   iv. Dedication – assist outsider in return for compensation
C. Objective: theft or sabotage on nuclear facility
D. Strategy: neutralization of security measures

An insider is categorized into three type A, B, and C with three path concepts; Intrusion, Guidance, and Attack. Intrusion affects the outer path such as (1 → 2), (1 ↔ 3), and (1 ↔ 4). Guidance covers some paths that are inside the limited area like (2 ↔ 3), (2 ↔ 6), and (4 ↔ 6). Attack covers two paths which interact the target directly such as (3 ↔ 5) and (6 ↔ 7). Type A includes only Intrusion paths. Type B includes Intrusion and Guidance paths. Type C includes Guidance and Attack paths.

It is assumed that insider type can affect both non-detection probability and travel time. The non-detection probability was assumed to increase by 10%, 15%, and 20% in type A, B, and C respectively. The travel time was assumed to decrease by 20%, 15%, and 10% in type A, B, and C respectively.

### *3.2. Analytic framework*

The two-person Stackelberg game is formulated by using a mixed integer programming (MIP). The actual model is constructed with the GAMS software program [5] that also returns expected consequence results.

### *3.3. Baseline problem*

Based on the multi-target mirrored network and the MIP formulation, we define the following baseline problem with zero budget (B=0) for security upgrades. The baseline network was assumed to have the default security measures defined previously. Hence, even in the absence of the security upgrades described below, an adversary is confronted with significant security measures. Numerical parameters in this section are assigned again for illustrative purposes.

### *3.4. Defender upgrades*

The security upgrades include both design changes and security measures that may be installed after construction. This example assumes a single budget for all upgrades, yet it can be modified to separate the upgrades into security by design and operational security categories with their own budgets. Summary of security upgrades are shown below.

Table III. Summary of upgrade cost and effect

| Upgrade ID | Cost | Impact |
|---|---|---|
| A | $$ | Reduce non-detection probability by 15% for arcs on either side of target. |
| B | $$$ | Reduce non-detection probability by 20% for all arcs on limited area |
| C | $ | Increase travel time by 25% for all arcs on bypass fence paths |
| D | $ | Reduce response time by 20 seconds. |
| E | $$$$ | Increase defeat probability from 0.7 (inside critical detection region 2, outside critical detection region 1) and 0.8 (outside critical detection region 2) to 0.85 and 0.9 respectively |

### *3.5. Insider threat*

Insider is an individual with authorized access to a facility and system who use their trusted position for unauthorized purposes. Insider is able to take advantage of their access rights and knowledge of a facility to bypass dedicated security measures. He/she can also capitalize on his/her knowledge to exploit any vulnerabilities in safety-related systems, with cyber security of safety-critical information technology systems offering an important example of the 3S interface. Because insider is capable of carrying out destructive actions not available to outsiders and have more opportunities to select the most vulnerable target and the best time to execute the malicious act, insider attacks are perhaps the key threat to the safety-security interface.

Insider is categorized by his/her working area in this model. Three path concepts cover exterior, intermediate, and interior area of the facility. Capabilities or security authorization level of Insider are different according to insider's workplace. Insider do not act solely without outsiders and he/she just assist by neutralizing relevant security measures. His/her assistance can raise non-detection probabilities and reduce travel times.

## 4. Results and Discussion

Game theoretic modeling is obviously inapplicable at a zero budget as there are no defender decisions to be made. With non-zero budgets, this model can be applied to choose the optimal security upgrades portfolio. Fig 3 shows a summary of expected consequences as a function of budget increase from zero to eleven for all types. And Tables IV and V summarize the upgrades that are purchased at each budget level. At all groups, the expected consequence values are decreasing with the budget level. And the expected consequence values increase from those of No insider type to Type A, B, C. Insiders marked as Type C have the largest capability to affect the non-detection probability. So the expected consequence of adversary is the highest at Type C. However, there is no significant difference between the capability of insider type A and B, and at certain budget level such as level 7, the expected consequence value of type A is even higher than that value of type B. Because the influence of some security upgrades affect sensitively more to the path of insider type B. Not only expected consequence but the solution of security upgrades also change in some budget levels. Security upgrades portfolio could also in consideration of the insider's assistance to adversary.

Table IV. Summary of results for each budget level; No insider and Type A

| | No insider | | | Type A | | |
|---|---|---|---|---|---|---|
| budget | Expected consequence | solution | target | Expected consequence | solution | target |
| 0 | 0.3909 | none | FCF | 0.4100 | none | FCF |
| 1 | 0.3537 | C | RSCS | 0.3731 | C | RSCS |
| 2 | 0.3537 | C | RSCS | 0.3731 | C | RSCS |
| 3 | 0.3247 | A,C | RSCS | 0.3412 | A,C | RSCS |
| 4 | 0.3148 | E | FCF | 0.3363 | E | FCF |
| 5 | 0.2980 | C,E | RSCS | 0.3198 | C,E | RSCS |
| 6 | 0.2826 | A,E | FCF | 0.3008 | A,E | FCF |
| 7 | 0.2653 | A,C,E | RSCS | 0.2838 | A,C,E | RSCS |
| 8 | 0.2653 | A,C,E | RSCS | 0.2838 | A,C,E | RSCS |
| 9 | 0.2461 | A,B,E | FCF | 0.2626 | A,B,E | RSCS |
| 10 | 0.2460 | A,B,C,E | FCF | 0.2626 | A,B,C,E | RSCS |
| 11 | 0.2460 | A,B,C,D,E | FCF | 0.2626 | A,B,C,D,E | RSCS |

Table V. Summary of results for each budget level; Type B and Type C

| | Type B | | | Type C | | |
|---|---|---|---|---|---|---|
| budget | Expected consequence | solution | target | Expected consequence | solution | target |
| 0 | 0.4196 | none | FCF | 0.4291 | none | FCF |
| 1 | 0.3699 | C | RSCS | 0.3925 | C | RSCS |
| 2 | 0.3699 | C | RSCS | 0.3925 | C | RSCS |
| 3 | 0.3384 | A,C | RSCS | 0.3576 | A,C | RSCS |
| 4 | 0.3336 | B,C | RSCS | 0.3576 | A,C | RSCS |
| 5 | 0.3162 | C,E | RSCS | 0.3416 | C,E | RSCS |
| 6 | 0.3075 | A,B,C | RSCS | 0.3191 | A,E | FCF |
| 7 | 0.2807 | A,C,E | RSCS | 0.3023 | A,C,E | RSCS |
| 8 | 0.2753 | B,C,E | RSCS | 0.3023 | A,C,E | RSCS |
| 9 | 0.2680 | A,B,E | FCF | 0.2792 | A,B,E | RSCS |
| 10 | 0.2460 | A,B,C,E | RSCS | 0.2792 | A,B,C,E | RSCS |
| 11 | 0.2460 | A,B,C,D,E | RSCS | 0.2792 | A,B,C,D,E | RSCS |

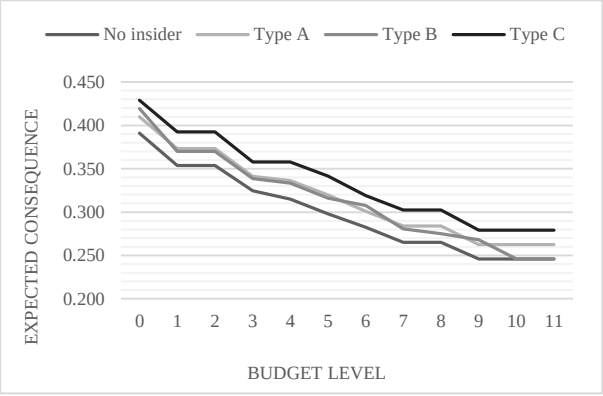


Fig 3. Summary of expected consequences as a function of budget

Fig 4 summarizes cost benefit analysis to an each insider type as a function of budget level. The cost-to-benefit ratio is higher at low budget level, for example, at level 1 and 3 for all types of an insider. The effectiveness of security upgrades decreases as budget rises. If the amount of budget is limited, then security upgrades with budget level 1 and 3 are highly recommended according to the model.

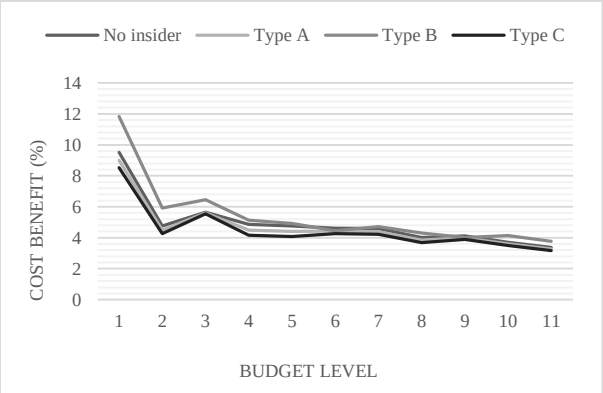


Fig 4. Cost-to-benefit ratio as a function of budget

## 5. Conclusion

This study revises and updates a game-theoretic model that was constructed for a physical protection system against attack by an intelligent adversary on a nuclear facility. Novel to our approach is the modeling of insider threat that affects the non-detection probability and travel time of an adversary. The game-theoretic approach has the advantage of modelling an intelligent adversary who has an intention and complete knowledge of the facility. In this study, we analyzed the expected adversarial path and security upgrades with a limited budget with insider threat modeled as increasing the non-detection probability. Our test case problem categorized three types of adversary paths assisted by the insider and derived the largest insider threat in terms of the budget for security upgrades.

More work needs to be done to incorporate complex dimensions of insider threats, which include but are not limited to: a more realistic mapping of insider threat, accounting for information asymmetry between the adversary, insider, and defenders, and assignment of more pragmatic parameter values. Considering the uncertainty and difficulty in obtaining the data for such parameters, uncertainty and sensitivity analysis will be needed in future work. To demonstrate the utility of the game-theoretic approach, a comparison between the use of game theory-based analysis and conventional security analysis will also be desirable.

## Acknowledgement

This work was supported by the Nuclear Safety Research Program through the Korea Radiation Safety Foundation (KORSAFe) and the Nuclear Safety and Security Commission (NSSC), Republic of Korea (Grant No. 1305017-0113-HD120)

## REFERENCES

[1] R. Ward, E. Schneider, A Game Theoretic Approach to Nuclear Safeguards Selection and Optimization, Dissertation at The Univ. of Texas at Austin, 2013
[2] L. A. Cox, Jr, Game Theory and Risk Analysis, Risk Analysis, Vol. 29, No. 8, 2009
[3] Proliferation Resistance and Physical Protection Evaluation Methodology Working Group, PR&PP Evaluation: ESFR Full System Case Study Final Report, Generation IV International Forum (GIF), 2009
[4] B. Canion, C. Hadlock, A. Zolan, D. Morton, J. E. Bickel, E. Schneider, Game-theoretic Allocation of Security Investments at Nuclear Reactors, Risk Analysis, under review, 2014.
[5] R. E. Rosenthal, GAMS – A User's Guide, GAMS Development Co., 2014