Advanced Reactor Safeguards & Security

Defensive Cybersecurity Architecture Design Using Force-on-Force Cyber-Physical Modeling

Prepared for
US Department of Energy

Lee T. Maccarone, Andrew S. Hahn, Benjamin R. Liu, Fred J. Oppel

Sandia National Laboratories

July 2025
SAND2025-09583R

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Currently, nuclear power plant physical security systems are highly dependent on air-gaps as a protective measure against cyber-threats. Cyber-physical threats become more likely as advanced cyber-threat capabilities to jump air-gaps transition into common use. Defending against the emerging threat of cyber-enabled physical intrusions is poorly understood. The consequence of these cyber-physical attacks has no quantitative analysis method to inform risk-informed, performance-based cybersecurity approaches. By modifying the physical security simulation tool Dante, cyber-physical threat consequence was able to be analyzed on a notional facility. The results of this analysis are used to design a Defensive Cybersecurity Architecture (DCSA) for the physical security system to produce example resilience measures for this notional facility. A DCSA defines security levels to provide a graded approach for defending plant functions, and security zones for trusted communication between systems. This approach can be applied to real world systems to produce physical protection systems and response measures that are resilient to cyber-physical threats.

ACKNOWLEDGEMENTS

This report was written for the Advanced Reactor Safeguards and Security program area in the U.S. Department of Energy Office of Nuclear Energy and satisfies milestone M3RS-25SN0402051. The authors would like to acknowledge the program leadership provided by Dan Warner (DOE-NE) and Ben Cipiti (Sandia National Laboratories).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

ACRONYMS AND TERMS

Acronym/Term	Definition
BBRE	Blast and Ballistic Rated Enclosure
CAS	Central Alarm Station
CDA	Critical Digital Asset
CONOPS	Concept of Operations
DBT	Design Basis Threat
DCSA	Defensive Cybersecurity Architecture
DFP	Defensive Fighting Position
DiD	Defense-in-depth
DTRA	Defense Threat Reduction Agency
ECP	Entry Control Point
IAEA	International Atomic Energy Agency
IT	Information Technology
NRC	Nuclear Regulatory Commission
NSS	Nuclear Security Series
OT	Operational Technology
PIDAS	Perimeter Intrusion Detection and Assessment System
PPS	Physical Protection System
RG	Regulatory Guide
RPG	Rocket-Propelled Grenade
SME	Subject Matter Expert
SMR	Small Modular Reactor
SNL	Sandia National Laboratories
SPO	Security Police Officer
TTX	Tabletop Exercise

This page left blank

1. INTRODUCTION

Physical security systems are essential to ensuring defense against design basis threats for nuclear power plants and are highly reliant on digital systems. These digital physical protection systems (PPSs) are assumed to be protected against cyberattacks through air-gapping their networks. The assumptions of the effectiveness of air-gaps were fundamentally broken by the Stuxnet attack in 2010 [1]. The Ramsay malware and ProjectSauron have demonstrated more recently that adversaries have the capabilities to jump the air-gap and are developing advanced methods to do so [2, 3]. Once the air-gap is defeated the networks and systems that constitute the PPS are relatively soft targets, often they are flat networks and do not benefit from timely updates and patches due to their isolated nature. Adversary access to these systems could allow total system defeat without contingency. Understanding this threat landscape and developing defensive methodologies based on empirical evidence is overdue by over a decade.

In 2023 Sandia National Laboratories (SNL) hosted the first large scale cyber-physical blended exercise to better understand the precipitation of physical attacks assisted by cyber adversaries [4]. When developing the attacks for this exercise, the cyber red team was able to gain control over the PPS once the air-gap was bridged using very simple cyber-attack methods (i.e. packet replay, network traffic sniffing, old opensource exploits). The air-gap was bridged using widely available, inexpensive commercial products which were installed by an insider. In under a week a small cyber red team of three was able to develop and deploy attacks to obfuscate cameras, control the biometric access system, and shutdown the PPS servers. Defenders struggled to cope with this threat due to the flat and low visibility architecture which is typical of a PPS. This exercise demonstrated that the difficultly of attacking this system was overestimated, the cyber risk of the PPS is poorly characterized, and that nuclear facilities are not prepared to handle cyber-physical threats. As state-level actor capabilities trickle into the mainstream cyber-attack toolset, nuclear facilities need to be provided defensive methods and tools which measurably improve security.

Nuclear control systems are facing similar issues with regards to cyber threats and the Defensive Cybersecurity Architecture (DCSA) concept offers an approach to passive cyber defense. Employing a risk-informed, performance-based DCSA design approach allows for the development of robust cybersecurity-by-design PPSs. The major issue is assessing the risks and robustness of a security system's design which relies significantly on human security personnel. Ideally, DCSA development is driven by deterministic cyber consequence analysis, the heavy reliance on human response and combat makes deterministic analysis of the PPS unfeasible.

To properly inform the DCSA process, each cyber driven failure of individual and combinations of PPS components must be evaluated. Current practice for PPS effectiveness analysis is rooted in table-tops and full-scale exercises, which are expensive and are considerably prone to biasing. A statistically significant analysis across even a small number of security system components would require hundreds of exercises. The answer to this dilemma is to simulate the facility, PPS, personnel, and cyber-physical threats virtually.

The SNL-developed Dante simulation platform provides the ability to fully simulate combat engagements, security systems, and defender and adversary behavior. With slight modifications, Dante can simulate cyber-attacks against the PPS allowing an analysis of the cyber risk profile. The cyber risk profile can then be used to inform the DCSA of a given nuclear power plant PPS. This report documents a cyber risk analysis of a notional small modular reactor (SMR) facility, which is used to inform an example DCSA of the PPS system.

It is important to note that the findings of this example analysis should not be used to derive direct system design or operational recommendations. The complex nature of the PPS and its interactions with personnel and response strategies requires each individual facility to undergo a full analysis. The process presented can be applied to a real facility, but the design of the building, the locations and types of sensors and cameras, and the response procedures will change the results of an analysis significantly.

2. DEFENSIVE CYBERSECURITY ARCHITECTURE BACKGROUND

The following overview of DCSA is quoted from [5]. For more information about DCSA, readers are encouraged to refer to [5]. The U.S. Nuclear Regulatory Commission (NRC) Regulatory Guide (RG) 5.71 states:

"An overall cybersecurity defensive strategy for a site must employ defense-in-depth strategies to protect CDAs [critical digital assets] from cyberattacks up to and including the DBT [design basis threat]. One acceptable method for achieving this goal is to incorporate a defensive architecture that establishes formal communication boundaries (or security levels) in which defensive measures are deployed to detect, prevent, delay, mitigate, and recover from cyberattacks. An example of such a defensive architecture is one that includes a series of concentric defensive levels of increasing security that conceptually correspond to existing physical security areas at a facility (e.g., vital area, protected area, owner-controlled area, corporate accessible area, public area)" [6].

The International Atomic Energy Agency (IAEA) defines the features of DCSA in the Nuclear Security Series (NSS) publication 17-T [7]. Several key definitions are quoted below from NSS 17-T.

- Facility Function: "a coordinated set of actions and processes that need to be performed at a nuclear facility" [7].
- Security Level: "a designation that indicates the degree of security protection required for a facility function and consequently for the system that performs that function" [7].
- System: "A set of components which interact according to a design so as to perform a specific (active) function, in which an element of the system can be another system, called a subsystem" [8].
- Security Zone: "a logical and/or physical grouping of digital assets that are assigned to the same computer security level and that share common computer security requirements owing to inherent properties of the systems or their connections to other systems" [7].

The relationships between these four elements are shown in Figure 2-1. Figure 2-1 depicts relationships common in existing fleet, leveraging the wrap-around approach. Security level requirements are shown as only related to zone boundaries, as system designs of existing fleet are unlikely to consider system changes for cybersecurity. However, the current design maturity of advanced reactor designs may allow for system design changes to simplify implementation and monitoring of cybersecurity as well as providing protection integrated within the system, unable to be bypassed by simple access to the internal areas of the zone.

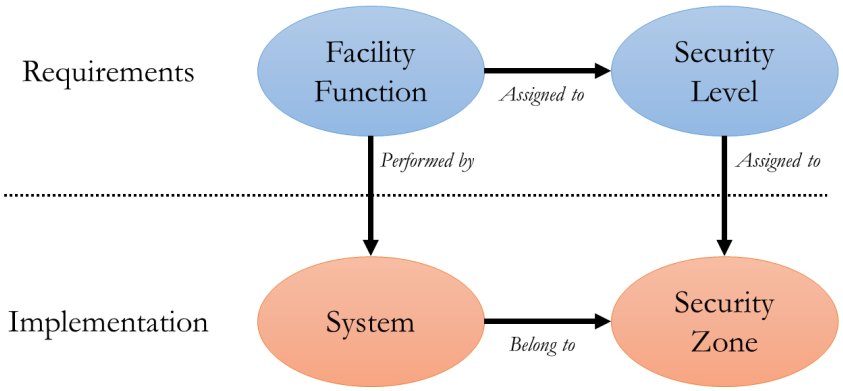

Figure 2-1. Relationship Between DCSA Elements (Adapted from [7])

A zone is a region bounded by logical and physical protections which contains at least one system. Communication between assets within a zone is trusted, while communication between different zones is restricted and controlled [7]. DCSA levels provide a framework for implementing a graded approach where security measures correspond to the significance of the functions assigned to each level. Each facility function is assigned a level based on its criticality. The stringency of measures put in place for a given level is directly related to the significance of the function protected by the level. Levels allow flexibility in security requirements across the facility which allows designers to prioritize the areas of greatest risk. Each level includes one or more zones. Zones enable defense-in-depth (DiD) if systems performing redundant functions are placed in separate zones. By placing systems performing redundant functions in separate zone, the adversary is forced to compromise multiple zones in order to prevent the function from being performed. Figure 2-2 provides an example of how DCSA zones and levels would be implemented. Note that Figure 2-2 shows the level nomenclature used by U.S. NRC; IAEA follows a nomenclature that ranges from security level 1 to 5, with security level 1 receiving the most stringent security requirements.

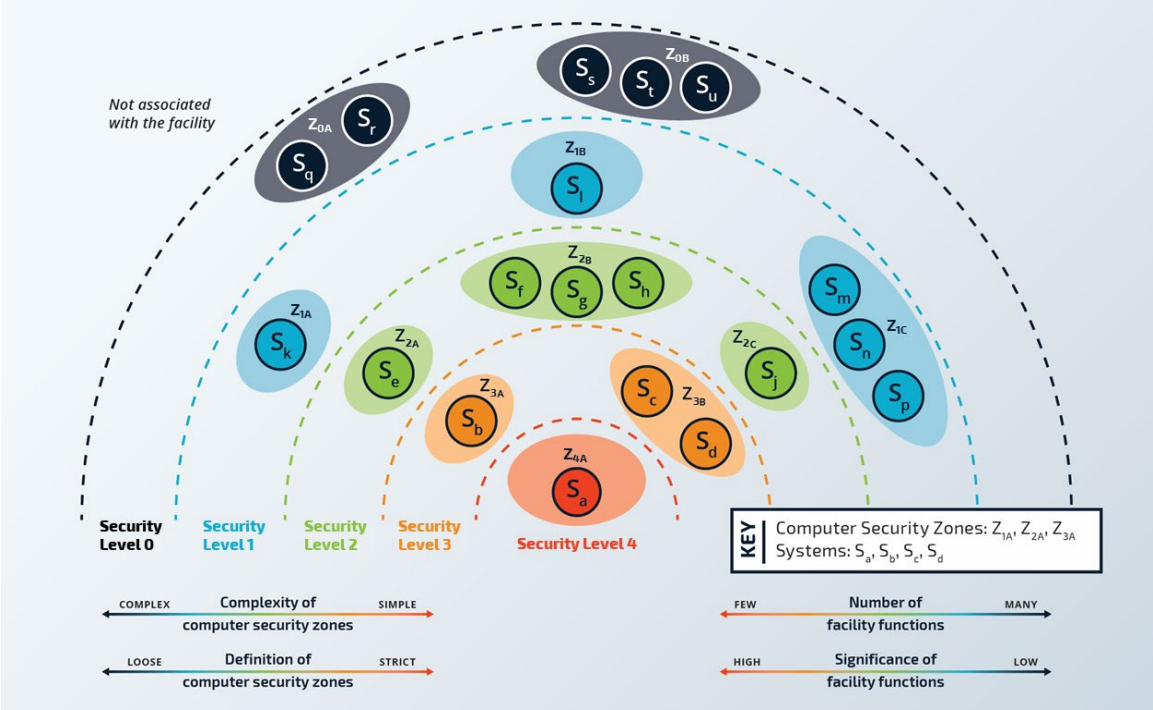

Figure 2-2. Conceptual DCSA Model [7]

One challenge in applying DCSA to PPS is the consideration of functional interdependencies. Within the context of PPS, the primary functions of concern are detection, delay, and response functions [9]. Detection must occur before delay can occur, and response requires a minimum amount of delay to be successful. These functions may have several interdependencies, for example, an information dependency exists when a response force requires the location of the adversary after detection occurs [7]. To scope this initial experiment, the detection function was examined. Future work should consider the impacts of cyber-attacks on digital PPS supporting detection, delay, and response functions.

3. EXPERIMENTAL SETUP

In this study, the Dante force-on-force modeling and simulation software was employed to analyze the effectiveness of various security measures against potential cyber and physical threats using a simulated small modular reactor (SMR) plant. The experimental setup was crafted to simulate real-world scenarios within a protected facility, providing a robust platform for evaluating the effectiveness of various security protocols. This simulation included detailed representations of critical elements such as building interiors, reactor components, and security assets, while ensuring the exclusion of any potentially sensitive information to maintain confidentiality and security standards. A realistic attack scenario was then developed which was informed by feedback and insights from subject matter experts (SMEs) with extensive experience in the field, allowing the creation of a credible threat landscape that could be effectively analyzed. This experimental setup aimed to showcase how DCSA could be utilized and studied further using simulation techniques.

3.1. Dante Software

Dante is a simulation tool developed by SNL to inform decision makers and response forces on the effectiveness of existing and future combat systems. Dante supports policy and concept of operations (CONOPS) development for national and local decision makers and first responders. Dante can be used to simulate force-on-force engagements and physical security system effectiveness. Analysts set up scenarios through the Dante Scenario Editor and execute multiple data runs using batch mode processing. Dante generates large quantities of stochastic data for assessments, trade-off studies, and sensitivity analyses about a given scenario. This data is used to discover and explore the outcomes or impacts from inserting technologies for force-on-force engagements or physical security systems. Dante is currently being used by several government agencies; the Defense Threat Reduction Agency (DTRA) has conducted a *verification and validation* effort to certify the use of Dante to analyze force-on-force engagements.

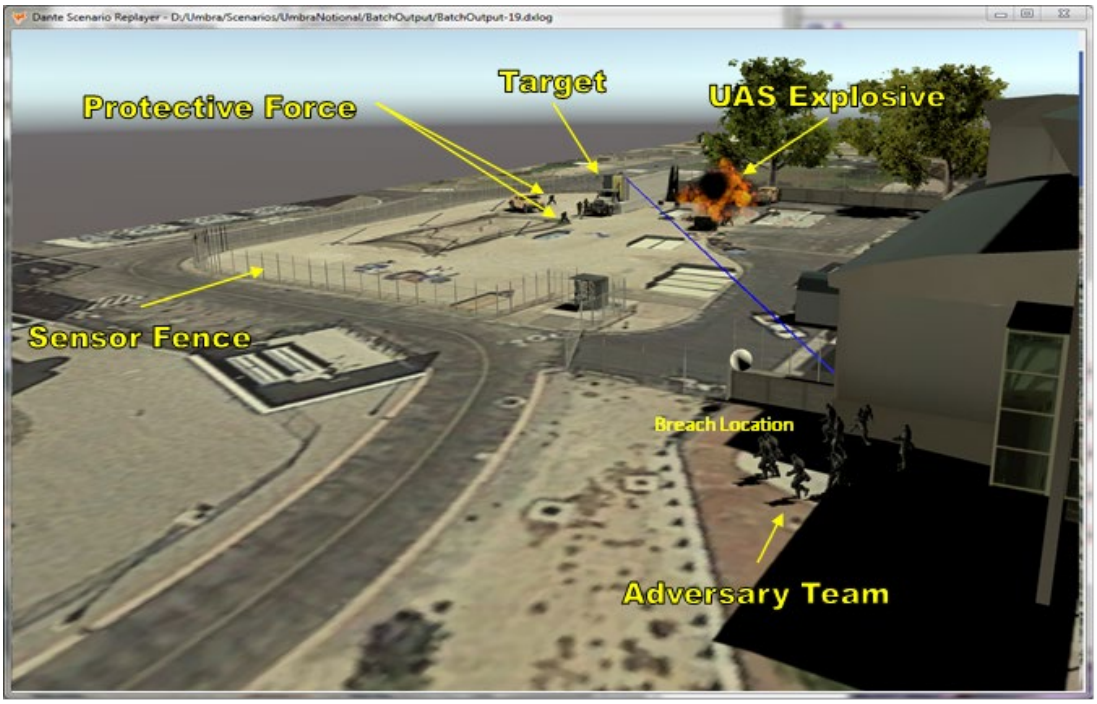

Figure 3-1: Example of a Dante scene with physical entities.

Dante scenarios include simulation of both physical assets as shown in Figure 3-1 above (e.g. buildings, people, sensors, weapons, etc.) and nonphysical elements (behaviors, communications, cyber systems). Scenario building in Dante is extremely flexible, allowing for variation in tactics, behaviors, procedures, perceptions, or locations of various assets. This level of simulation allows Dante to run attack scenarios fully autonomously, removing one of the key roadblocks to robust data collection and analysis – cost. Cyber-physical attacks of this scale traditionally require tabletop exercises (TTX) to be conducted, which require large numbers of participants and their respective number of hours of effort. A single simulation run can produce a similar output to a fully staffed TTX for a fraction of the cost and effort. In addition, Dante can run Monte Carlo batch simulations that produce large amounts of data, allowing analysts to explore the statistical results of attacks against the site's security systems.

For our purposes, the Dante simulation environment was extended to include cyber affect models. A Dante cyber world module was created to enable cyber components to interact or influence physical security components. These interactions can occur from a scripted timeline of events based on experimental data or from a person injecting cyber events interactively such as a TTX environment.

3.2. Site Description

This experiment required a site that was a representative model of a nuclear reactor facility, including building interiors, reactor components, and security assets, while avoiding the inclusion of any potentially sensitive information. An SNL-designed facility for a generic small modular reactor (SMR) plant was selected and turned into a detailed 3D model for use in Dante. The facility is entirely notional, although real-world reactor and facility concepts were used for inspiration [10]. The complete layout described in the SNL design document (Figure 3-2) is nearly identical to the model created for Dante (Figure 3-3), albeit with minor tweaks to the locations of certain buildings, doors, and offsite entities. Facility sensor and zone configurations are based on real-world plant design, and both red and blue team strategies are based off a combination of previous TTX experience and SME recommendations for best practices.

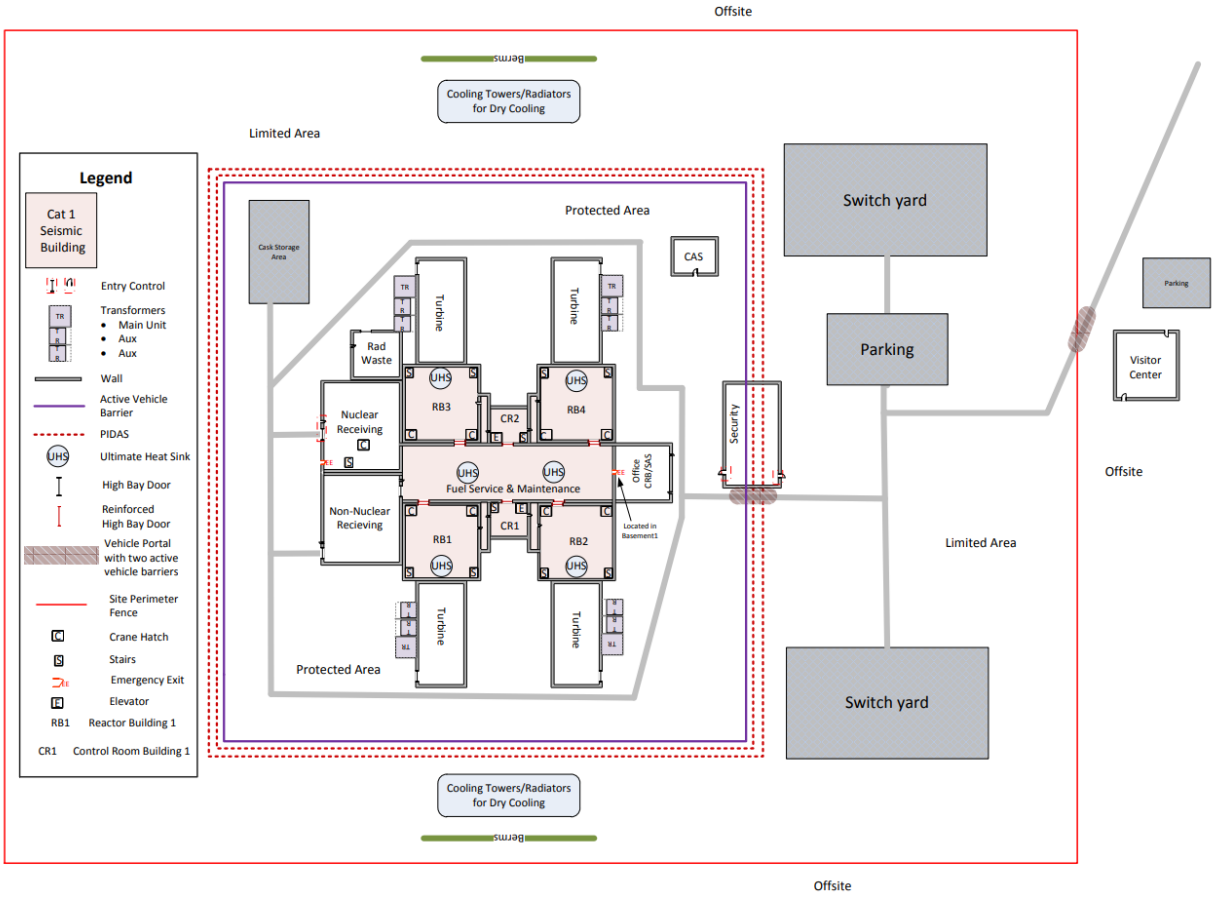

Figure 3-2: SNL Generic SMR Design [10].

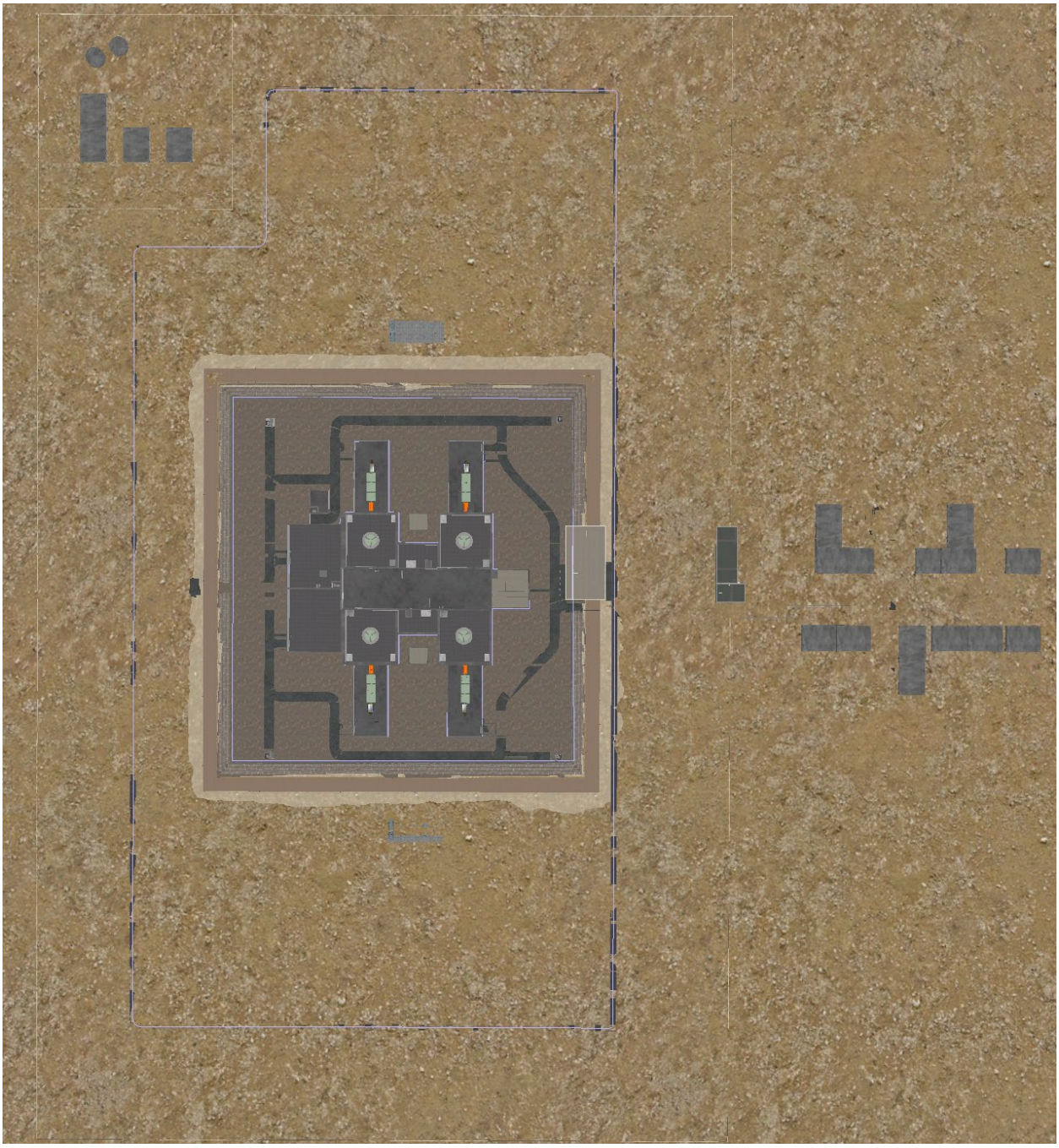

Figure 3-3: Simulated 3D model in Dante.

3.3. Adversary Attack Profile

A formal Design Basis Threat (DBT) was not considered due to sensitivity concerns; however, a realistic attack profile was designed using SME feedback and experience. The attack consists of a total of six (6) attackers, well equipped with weapons and explosives. [See Figure 3-4] Three attackers are equipped with 7.62mm assault rifles and rocket-propelled grenades (RPGs), and three are equipped with 5.56mm assault rifles and assorted breaching equipment.

The scenario assumes that this attacking force is highly motivated, has military training and skills, and that each individual has the willingness to kill and be killed. Additionally, it is assumed that the attack planners have the knowledge to identify specific equipment and locations to plot a successful attack. Although not strictly modeled, the scenario assumes an additional external adversary that is conducting cyber-attacks on information technology (IT) and operational technology (OT) systems specifically on the PPS.

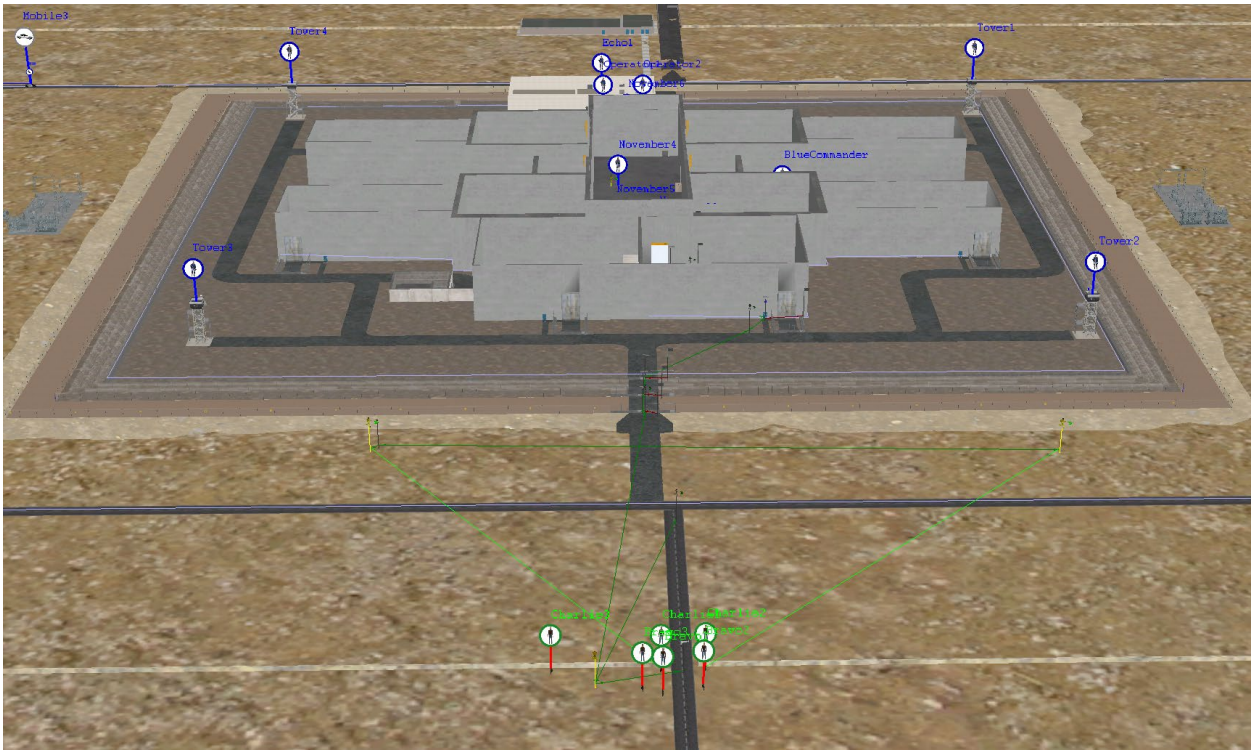

Figure 3-4: Initial attack location on facility.

Given these assumptions, an attack strategy was formed that maximizes the red team's chances of success. Due to the site's layout the west side is the least heavily defended, with only two tower guards and a mobile unit potentially engaging the adversary before they reach the building. This attack vector also provides relatively easy access to the reactor buildings.

The attack starts at the west side of the facility with all six attackers gathered outside of the designated limited area. They breach the perimeter fences and move into the building, continuing to breach doors along the way, with the ultimate goal of breaching the reactor housing and causing a radiological release. For the sake of simplicity, we consider a breach of the containment housing hatch (at ground level) a successful attack. Figure 3-5 shows the general attack plan.

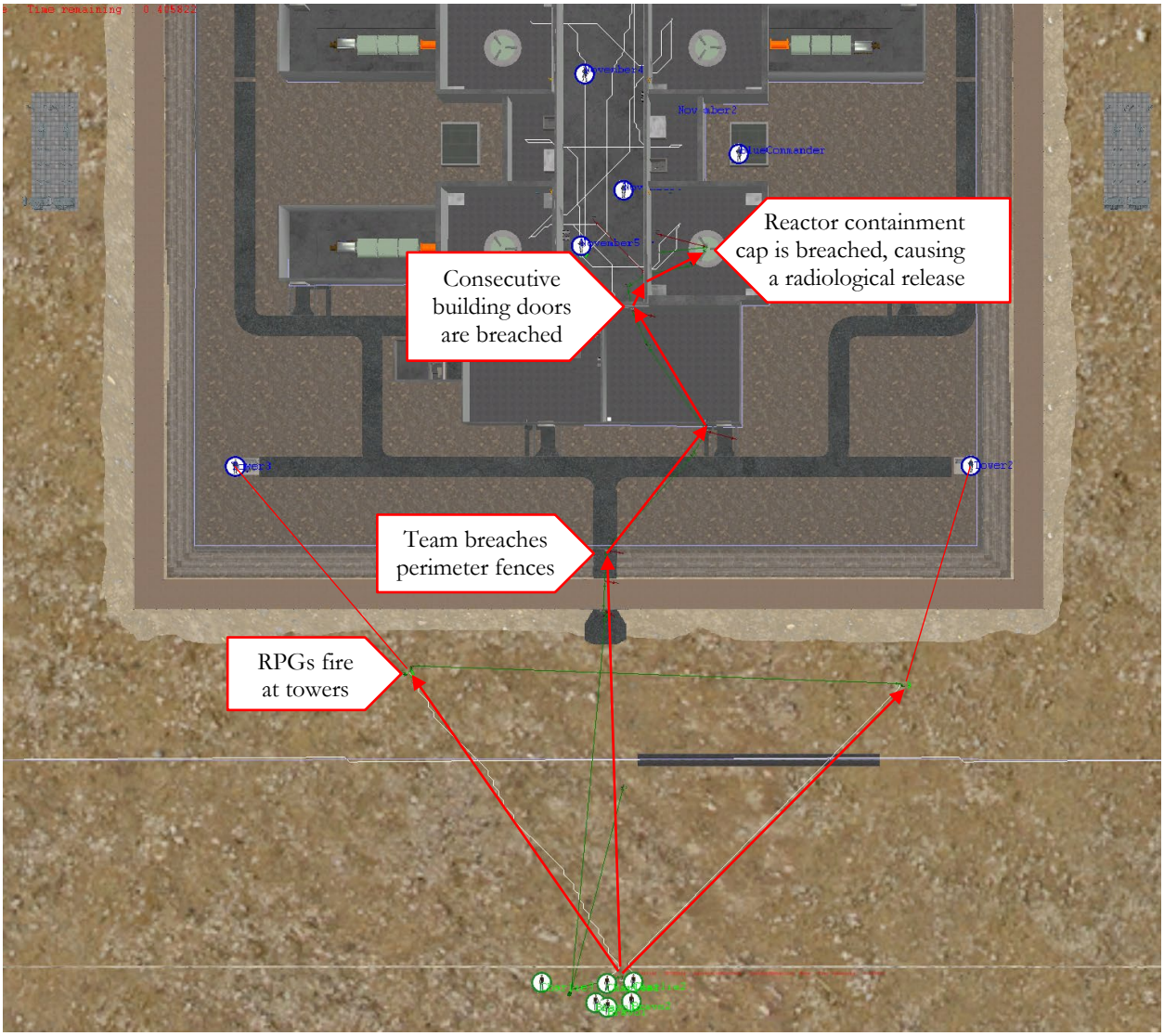

Figure 3-5: Red team attack strategy.

3.4. Security Personnel and Equipment

Blue team personnel and facility sensors are set up in a notional high-security configuration. There are a total of twelve (12) armed security police officers (SPOs) -- two (2) in a mobile unit that patrols the perimeter in a police car, four (4) in Blast and Ballistic Rated Enclosure (BBRE) towers, and six (6) patrolling the reactor building interior. Two Central Alarm Station (CAS) operators, an entry control point (ECP) operator, and a blue team commander are also modelled but have no physical effect on the simulation. The building guards, labeled in Figure 3-6 as *November1 – November6*, are each assigned a different level of the facility to patrol. Each have the ability to investigate and respond to potential threats within the interior of the building. All armed guards are equipped with a 5.56mm assault rifle and radio communication devices connected to the CAS.

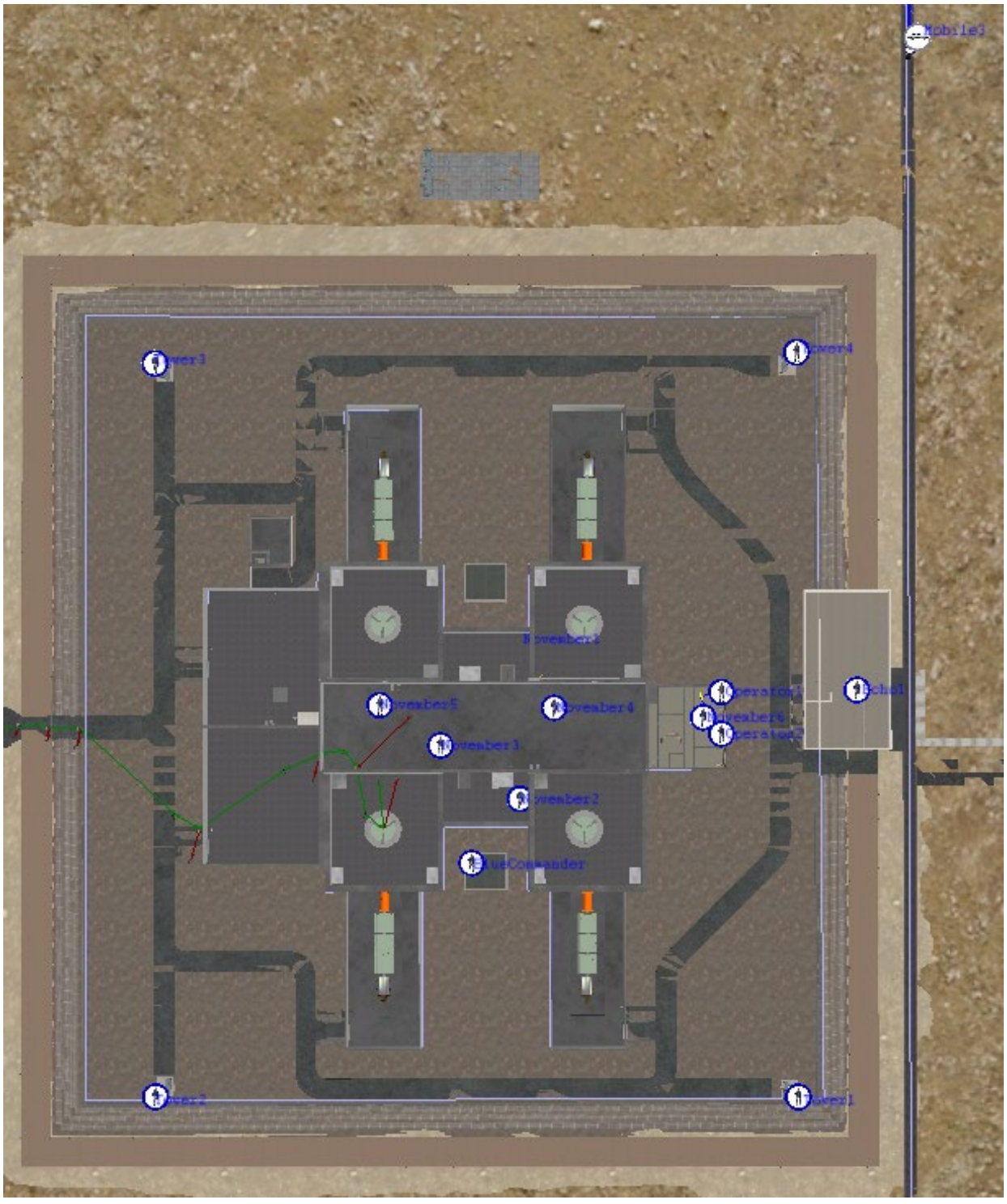

Figure 3-6: Blue team personnel.

Facility sensors are then divided up into categories:

- Fence/perimeter sensors, which include an industry-standard Perimeter Intrusion Detection and Assessment System (PIDAS) sensor array located at the second fence line and cameras placed atop the BBRE towers. Although this comprises of two technically different systems,

considering them as a unified system is necessary to have both detection and assessment capabilities for threats at the fence perimeter. In the context of DCSA considerations, the network topology places both of these sensor systems in the same group.

- Building cameras, with the relevant cameras placed at the building exterior and inside the reactor containment room.
- Door sensors, at each exterior and interior door to detect unauthorized intrusion.
- Motion detectors, covering the entirety of each receiving bay.

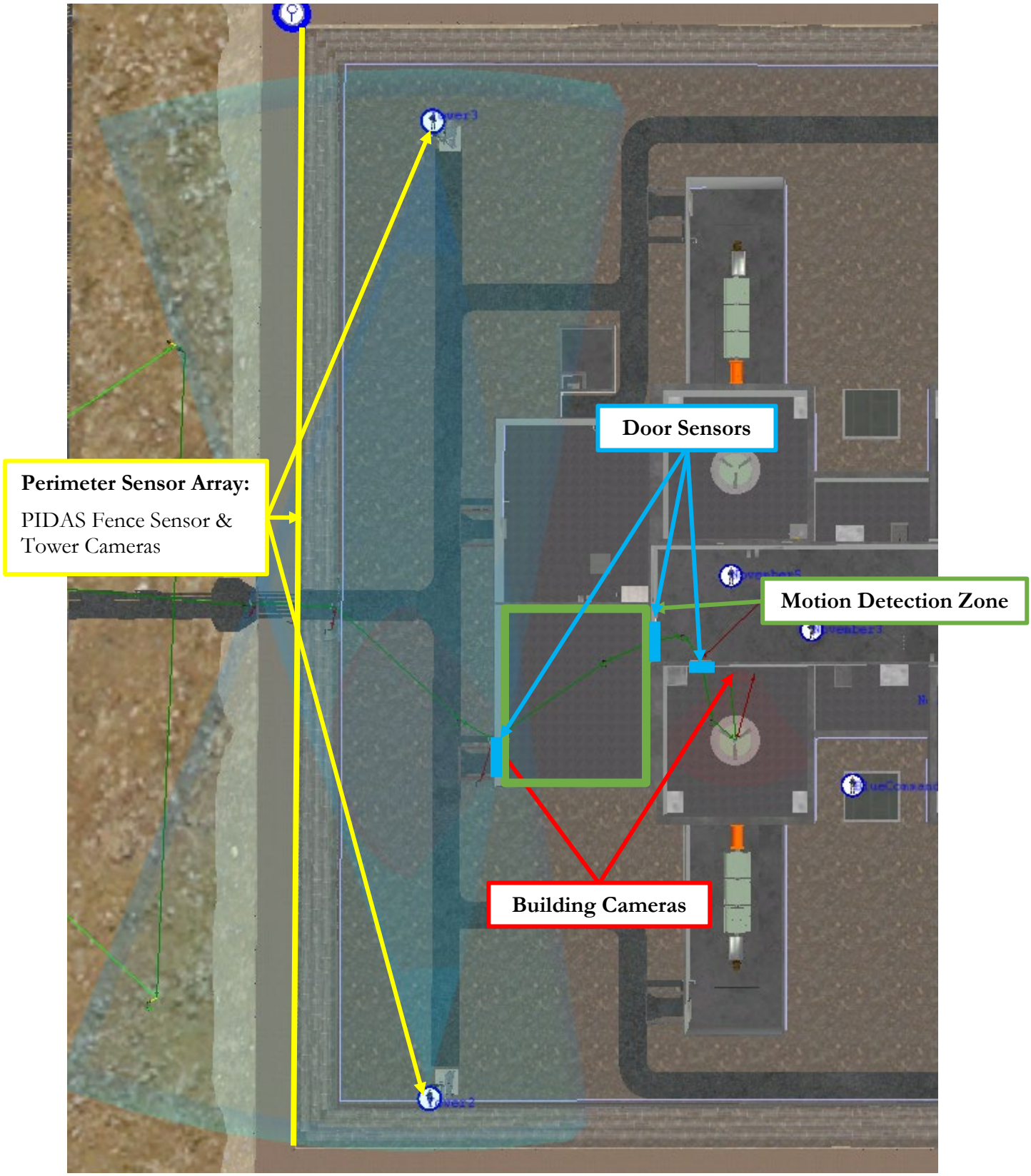

Figure 3-7: Facility sensor placement.

Responses to sensor alarms are handled by the *Blue Commander*, a special entity that takes sensor readings from the CAS and commands blue team members to respond appropriately. To simulate real-world conditions and cover potential human element impacts, sensor reports are sent through a chain of command with predetermined delays at each step. Every sensor hit (including human reports) gets sent to the CAS first. The CAS has a 10 second initial "process delay" upon first contact, which then drops to 3 seconds for subsequent reports. Once processed, the Commander must decide on a plan of action and command the blue team to respond. The commander then has a "decision time" of 5 seconds, after which commands are confirmed by individual SPOs within the next 10 seconds.

This scenario includes two key assumptions: first, the assumption that the communication chain from SPO – CAS – Commander is not disrupted and there is perfect communication across the board. Simulating communications failure/disruption does exist as a feature in Dante, but for the sake of simplicity was excluded from the scenario. The other assumption is that defenders are not aware that they are under cyber-attack and make no operational changes in response. Despite its possibility in real-world scenarios, there does not exist a sufficient field of evidence to assume any kind of behavioral changes based on knowledge of an active cyber-attack.

In Dante, there are several response commands that the commander can execute: Investigate, Contain, and Secure. Investigate commands are triggered by detection sensor hits that are not immediately assessable by other sensors to confirm a threat. This includes human sensing and responses to events like visual contact and audio cues (gunshots, breaching, etc.) In this scenario, this also includes alarms sent by the PIDAS sensor, door sensors, and motion detectors. In the case of an investigate command, one or more team members are sent to the area to attempt assessment of the threat.

Containment commands are responses to confirmed high-level threats. These events are triggered by any sensor with assessment capabilities, in this scenario exemplified by the tower and building cameras. When a containment command is sent out, blue team members go to predesignated defensive positions within the building to attempt to stop the adversary from reaching their target.

Secure commands indicate that an asset is in imminent danger and tell all available personnel to go to and defend the area as quickly as possible. In this scenario, the only example is triggered by the camera inside of the containment hatch room. Once adversaries are detected and assessed to be inside that room, all available building guards immediately move in to stop a potential breach.

Figure 3-8: Fuel service and maintenance hallway between receiving bay and reactor containment room. Note the Defensive Fighting Position (DFP) on the right – these are utilized when containment orders are given.

Each type of command is given a certain level of priority in ascending order (0, 1, 2, ...), where higher values both override lower ones and indicate a higher level of response. This set of commands and priority levels is how the entirety of the site's security policy is defined. Table 3-1 shows all relevant sensors, their command responses, and priority levels. Note that the type of sensor also plays a role in how blue team responds. Detection alone triggers an investigatory response, while having positive assessment of a threat allows the commander to take more drastic containment measures.

Additionally, SPOs have simulated visual and auditory sensing and can conduct their own threat analysis. Each SPO is equipped with a radio and has the ability to send contact reports to the CAS directly; contact reports are formatted in a similar fashion to facility sensor reports. However, for this particular scenario and the security policy chosen for this site SPO contact reports are considered purely a detection measure. A contact report indicating that a threat may exist will trigger the commander to send additional personnel to investigate. However, a report of audible gunfire or of the officer being under fire does increase the priority of the investigative measure.

Table 3-1 Facility sensor response and priority.

Sensor Description [Zone]	Sensing Type	Response	Priority
Contact Report **[Personnel]**	Visual/Auditory, Detection only	Investigate	1 (4 if gunfire is detected)
PIDAS Fence [**Fence Sensors**]	Motion, Detection only	Investigate (Fence)	3
Tower Cameras [**Fence Sensors**]	Visual, Assessment only	Contain (Building Interior)	5
Building Exterior Camera [**Cameras**]	Visual/Motion, Detection and Assessment	Contain (Building Interior)	5
Receiving Bay Detector [**Motion Detector**]	Motion, Detection only	Investigate (Receiving Bay)	15
Receiving Bay Door Sensor [**Door Sensors**]	Motion, Detection and Assessment	Contain (Receiving Bay)	10
Hallway Door Sensor [**Door Sensors**]	Motion, Detection and Assessment	Contain (Hallway)	20
Containment Room Door Sensor [**Door Sensors**]	Motion, Detection and Assessment	Contain (Containment Room)	30
Containment Room Camera [**Cameras**]	Visual/Motion, Detection and Assessment	Secure (Containment Room)	40

4. RESULTS AND ANALYSIS

This section provides the simulation results, statistical analysis, and DCSA design for the previously described experiment. First, the raw data is examined and general trends are identified. Next, the data is analyzed using statistical methods to determine whether the trends are statistically significant and whether any additional conclusions can be drawn. Finally, the statistical analyses are leveraged to design a DCSA for this PPS. These results are based on the specific site analyzed in these experiments and the results may not be generalized to every site.

4.1. Simulation Results

Table 4-1 summarizes the simulations conducted in this experiment. Each simulation case is described by the set of digital systems that are cyber-attacked for those simulations. For notational convenience, we assign variable names to each of the digital systems (i.e., X1, X2, X3, and X4). Four digital systems were considered, therefore a total of 16 simulation cases were identified corresponding to each of the possible combinations of compromised digital systems. Case 1 corresponds to the scenario where all four digital systems are operating normally, and Case 16 corresponds to the scenario where all four digital systems are compromised.

Table 4-1. Simulation Cases Overview

Digital System Attacked	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Door Sensors (X1)		X				X	X	X				X	X	X		X
Cameras (X2)			X			X			X	X		X	X		X	X
Motion Detector (X3)				X			X		X		X	X		X	X	X
Fence Sensors (X4)					X			X		X	X		X	X	X	X

A total of 600 simulations were conducted for each of the cases summarized in Table 4-1. The remainder of this subsection summarizes the raw simulation outcomes. These simulation outcomes include Blue's win counts and the casualties suffered by Blue.

Blue's total wins for each simulation case are shown in Figure 4-1. Blue's baseline wins when all digital systems are operating normally is high (97.8%) and when the adversary compromises all digital systems, Blue's success rate is low (16.2%). When individual digital systems are operating normally (i.e., when the adversary comprises the other three systems), the fence sensor system (Case 13) and the camera system (Case 14) provide Blue with win rates near 100%. The door sensor system (Case 15) provides a notable improvement over the worst-case cyber-attack with a win rate of 64.8%. Notably, the motion detector system (Case 13) does not provide Blue with an advantage over the worst-case cyber-attack. Nearly all combinations of at least two systems operating normally provide Blue with win rates near 100%, except the combination of door sensors and the motion detector (Case 10). Case 10 performs similarly to Case 15 where only the door sensors are operating normally. The other two-system cases including the motion detector and another digital system also perform similarly to the cases where only the additional digital system is operating normally (Cases 8 & 13 for the cameras, and Cases 6 & 12 for the fence sensors). These results suggest that the motion detector does not improve Blue's likelihood of success for this attack scenario.

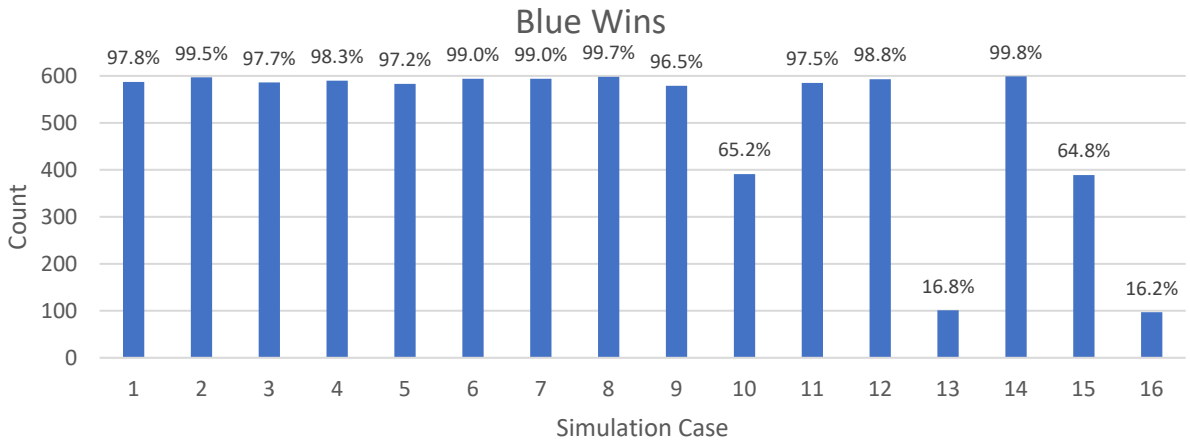

Figure 4-1. Number of Blue Wins for Each Simulation Case

Blue's casualties were examined to determine whether some digital systems contribute more significantly to Blue safely neutralizing Red. A heatmap of Blue's total casualties for each simulation case is shown in Figure 4-2. This heatmap implies that Blue's casualties are correlated with Red's win rates, as expected. Blue's casualties are similar across simulation cases with near-100% win rates for Blue, and those casualties appear to be significantly less than the simulation cases with lesser Blue win rates (Cases 10, 13, 15, and 16). Interestingly, Cases 10 and 15 (door sensors active) appear to have Blue casualties skewed slightly greater relative to Case 16, despite having greater Blue win rates in those cases than in Case 16. This may be because Blue is drawn to a more successful but less efficient fighting position when responding to the door sensors. Given that Blue's casualties appear highly correlated with Red's win rates, these casualties will not be investigated with statistical methods unless the analysis of Blue's win rates is insufficient.

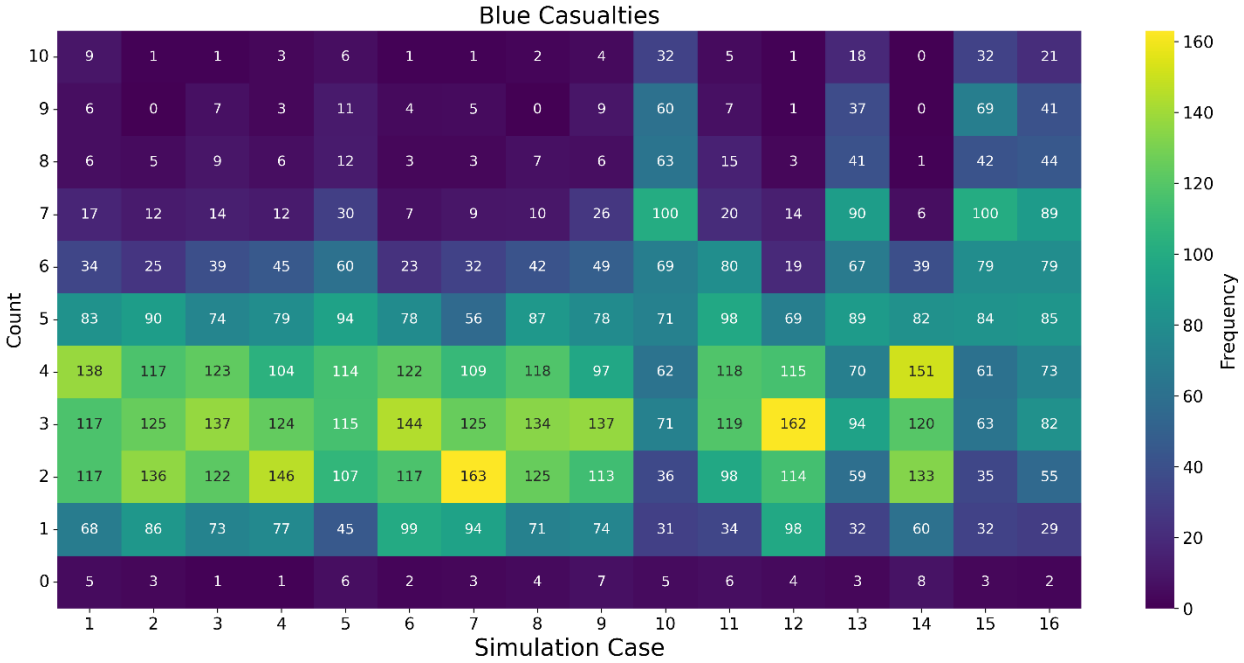

Figure 4-2. Heatmap of Blue Casualties for Each Simulation Case

4.2. Statistical Analysis

Logistic regression was used to analyze Blue's win rates shown in Figure 4-1. Logistic regression was selected because the outcome variable is binary (Blue either wins or loses) and because it enables us to analyze the effects of both individual predictor variables and their interactions [11]. The results of the logistic regression analysis are summarized in Table 4-2. The McFadden pseudo R^2 value is 0.555, indicating an excellent fit of the logistic regression model to the data (note that this McFadden pseudo R^2 value is approximately equivalent to an R^2 value of 0.9 as used in ordinary least squares statistical methods [12]).

The log-odds coefficient and odds ratio coefficient columns indicate the relationship between the predictor variable and the outcome variable. The log-odds coefficient shows the change in the log-odds of Blue's success when the predictor security system is compromised. The odds ratio coefficient is the exponentiated log-odds coefficient – transformed for ease of interpretation. The odds ratio coefficient gives the change in odds of success for a unit change in the predictor variable, when all other predictors are held constant. Note that odds are distinct from probabilities, and calculating the change in probability given the odds ratio coefficient requires the initial probability. A negative log-odds coefficient or odds ratio coefficient less than one indicates that the log-odds of Blue's success decreases when the security system is compromised. The interpretation of these coefficients is simple for the main effects (i.e., X1, X2, X3, and X4) but differs slightly for the interaction terms (e.g., X1:X2, X1:X2:X4). The coefficient for interaction terms represents the non-additive effect of all of the corresponding systems being compromised simultaneously. A negative log-odds interaction coefficient or odds ratio coefficient less than one indicates that the effect of compromising the set of systems is worse than the sum of the individual effects.

The columns z and $\mathrm{P}(>|\mathrm{z}|)$ correspond to the test statistic and the probability that the magnitude of the test statistic exceeding that value (p-value). Probabilities of less than 0.05 are statistically significant. The intercept (no cyber-attack), X1 (door sensor cyber-attack), X2:X4 (cameras and fence sensor cyber-attack interaction), X1:X2:X4 (door sensor, cameras, and fence sensor cyber-attack interaction) terms had statistically significant p-values, meaning that these terms are very likely to affect Blue's success rate.

Table 4-2. Logistic Regression Results

Terms	Log-Odds Coefficient	Odds Ratio Coefficient	Std. Error	z	P(>|z|)
Intercept	3.81	45.2	0.280	13.6	4.73E-42
X1	1.48	4.41	0.643	2.31	2.11E-2
X2	-0.0758	0.926	0.390	-0.195	8.46E-1
X3	0.268	1.31	0.425	0.630	5.29E-1
X4	-0.275	0.759	0.373	-0.737	4.61E-1
X1:X2	-0.622	0.537	0.809	-0.769	4.42E-1
X1:X3	-0.966	0.381	0.827	-1.17	2.43E-1
X1:X4	0.682	1.98	0.988	0.691	4.90E-1
X2:X3	-0.685	0.504	0.550	-1.25	2.13E-1
X2:X4	-2.83	0.0588	0.469	-6.04	1.50E-9

Terms	Log-Odds Coefficient	Odds Ratio Coefficient	Std. Error	z	P(>|z|)
X3:X4	-0.139	0.870	0.556	-0.250	8.03E-1
X1:X2:X3	1.23	3.41	1.06	1.16	2.46E-1
X1:X2:X4	-3.77	0.0231	1.11	-3.39	7.05E-4
X1:X3:X4	1.53	4.63	1.52	1.01	3.14E-1
X2:X3:X4	0.542	1.72	0.668	0.811	4.17E-1
X1:X2:X3:X4	-1.83	0.161	1.67	-1.09	2.74E-1

4.3. Defensive Cybersecurity Architecture Design

This work is primarily concerned with the assignment of systems to security zones given the system's importance to the security of the facility. For examples of the assignment of cybersecurity controls in a graded approach according to security level requirements, please refer to [5]. The following is a performance-based approach based on the logistic regression analysis performed on the Dante simulation data. To obtain DCSA design constraints, we will begin with individual predictors and progress to increasingly complex interactions as needed. We will examine the odds ratio coefficients to determine how a cyber-attack targeting a set of systems affects Blue's win rates.

The cameras and fence sensors (X2 and X4, respectively) had odds ratio coefficients less than one, therefore when those individual systems are compromised, it is expected that Blue's performance will be negatively affected. It is noteworthy that the door sensors and motion detector (X1 and X3, respectively) have odds ratio coefficients greater than one, indicating that when those individual systems are compromised it is expected that Blue's performance will be enhanced. This was an unexpected result (particularly the magnitude of the X1 coefficient) that may be explained by the priorities and modeling assumptions defining Blue's behavior in Dante simulations. Triggering the door sensors was modeled as the highest priority for Blue, resulting in Blue rushing to contain the threat and exposing themselves to Red. It is recommended that Blue's techniques be re-evaluated to determine if the modeled priorities and behaviors are optimal. We will proceed with the DCSA design with the available data, however updated data under new behaviors could be used to update the DCSA. Given this analysis, the cameras and fence sensors are more critical to the security of the plant than the door sensors and motion detector, and should be governed by more stringent cybersecurity requirements. In other words, although all of these systems perform detection functions, some are more critical to the security of the plant because of their implementation, and therefore those specific functions should be governed by cybersecurity requirements belonging to a higher security level.

Second-order interactions are evaluated based on both the coefficients of the individual terms and the interaction term. The "overall" odds ratio coefficient is the product of the coefficients of the two individual terms and that of the interaction term. Several second-order interactions have odds ratio coefficients of less than one, indicating that the combined effects of compromising both systems are worse for Blue than the additive effects of the compromise of individual systems. These considerations are summarized below:

- If two systems that perform functions governed by different security levels are to be assigned to the same security zone, the system performing the less critical function must be governed by the security level assigned to the more critical function performed by the other

system. We have already identified that the functions performed by the cameras and fence sensors (X2 and X4) ought to be protected at a higher security level than the door sensors and motion detector (X1 and X3) and we will assume that it is not desired to elevate a security system to a higher degree of protection than is required. Therefore, we will only consider the interaction terms X2:X4 and X1:X3 to determine if there are any constraints preventing them from being assigned to the same zones within their respective security levels.

- The compromise of the door sensors (X1) and the motion detector (X3) systems does not negatively affect Blue's win rate, despite the fact the odds ratio coefficient X1:X3 is less than one. This is because of the size of the odds ratio coefficient of X1. Given these results, we do not impose any constraints on whether the door sensors and motion detector may be placed in the same security zone.
- The compromise of the cameras (X2) and fence sensors (X4) significantly negatively affects Blue's win rate. The individual odds ratio coefficients of X2 and X4 were less than one and the coefficient of the interaction term was much less than one. Notably, the X2:X4 interaction term was statistically significant. Given these results, we constrain that the cameras and fence sensors shall not be placed in the same security zone.

Analysis of higher-order interactions is not necessary because we do not have a set of three systems that perform functions governed by the same security level. If we were to analyze higher-order interactions, we would use a method similar to that applied to second-order interactions, but we would consider all of the lower-order interactions preceding the higher-order interaction. For example, when considering third-order interactions, we multiply the odds ratio coefficients of the individual terms, those of the second-order interactions, and those of the third-order interactions. Most third-order interactions have odds ratio coefficients of greater than one, indicating that the combined effects of compromising the three systems are better for Blue than the additive effects of the compromise of individual systems.

Given this analysis, Figure 4-3 shows a DCSA design candidate for the four systems under consideration. The cameras and fence sensors are governed by the higher security level and are placed in separate security zones because of our performance-based design constraint. The door sensors and motion sensors are governed by the lower security level and are placed in the same security zone because there was no negative impact to Blue's win rate when both systems were compromised. For examples of how cybersecurity controls may be applied within the context of this DCSA, readers are encouraged to refer to [5].

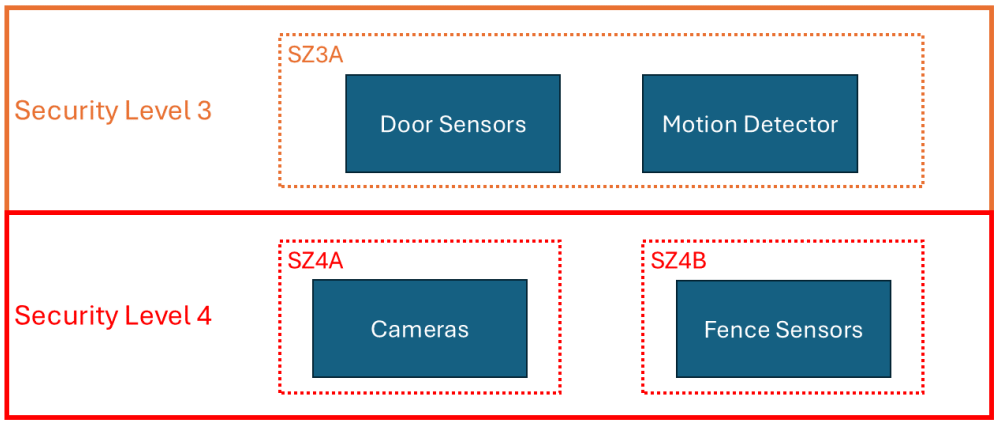

Figure 4-3: PPS DCSA Design

This page left blank

5. CONCLUSION

The analysis of a notional advanced reactor facility given in this report has demonstrated that it is feasible to perform performance-based DCSA design for a physical security system. The risk determination needs of the DCSA process can be satisfied by simulations powered by the Dante simulation software. Additionally, the response strategies of the defensive force can be analyzed under the pressures of a cyber threat. Designing the PPS system architecture and response strategies with this detailed cyber-physical analysis can create systems which are highly resilient against performance degradation due to cyber-attacks. Knowing which systems are the greatest contributors to security provides the basis to effectively and economically implement cybersecurity defenses and segment networks.

The demonstrated analysis can be performed on a real facility and create resilient defense-in-depth physical protection systems, but improvements with modeling can increase design assurance and response procedure effectiveness. This analysis assumes that the defenders have no knowledge of the cyber-attack on the facility. This could create some gaps in response analysis and result in overprotective assumptions. Defenders may not respond in the same way if they had knowledge that the system was under cyber-attack. They could become hyper vigilant or aggressive with their responses to alarms or delay their response until they are sure the alarm was not a distraction. Currently, the agents in Dante do not have the behavior models for cyber threat awareness because there is a lack of data for human responses to cyber-physical threat information. Additional research is needed to ensure that human responses to cyber threats are captured, but the DCSA-PPS approach demonstrated can still drastically improve the cyber-physical threat defense of advanced reactors.

This page left blank

REFERENCES

[1] N. Falliere, L. O. Murchu and E. Chien, "W32.Stuxnet Dossier," Symantec, 2010.

[2] ESET, "New cyber espionage framework named Ramsay discovered by ESET Research," 13 May 2020. [Online]. Available: https://www.eset.com/us/about/newsroom/press-releases/new-cyber-espionage-framework-named-ramsay-discovered-by-eset-research/?srsltid=AfmBOoqklW_4hc_2xwxj3EnrxstBLy12wlmEeqVfC2OfFS-L4-6guV6X. [Accessed 30 July 2025].

[3] Kaspersky, "The ProjectSauron APT," Kaspersky, 2016.

[4] M. K. Erdman, M. T. Rowland, A. S. Hahn, R. Pierce and A. M. Romero, "Canada-US Blended Cyber-Physical Security Exercise: Final Report," Sandia National Laboratories, 2023.

[5] L. T. Maccarone, M. T. Rowland, R. J. Brulles and A. S. Hahn, "Design of Defensive Cybersecurity Architectures for High Temperature, Gas-Cooled Reactors," Sandia National Laboratories, Albuquerque, NM, 2024.

[6] U.S. Nuclear Regulatory Commission, "Regulatory Guide 5.71 - Cyber Security Programs for Nuclear Facilities," Rockville, MD, 2010.

[7] International Atomic Energy Agency, "NSS 17-T: Computer Security Techniques for Nuclear Facilities," IAEA, Vienna, Austria, 2021.

[8] International Atomic Energy Agency, "IAEA Nuclear Safety and Security Glossary," IAEA, Vienna, Austria, 2022.

[9] M. L. Garcia, The Design and Evaluation of Physical Protection Systems, Burlington, MA: Butterworth-Heinemann, 2008.

[10] T. G. Lewis, B. B. Cipiti, S. E. Jordan and G. A. Baum, "Generic Small Modular Reactor Plant Design," Sandia National Laboratories, 2012.

[11] D. W. Hosmer, S. Lemeshow and R. X. Sturdivant, Applied Logistic Regression, Hoboken, NJ: John Wiley and Sons, 2013.

[12] D. McFadden, "Statistical Estimation of Choice Probability Functions," in *Urban Travel Demand: A Behavioral Analysis*, New York, NY, American Elsevier Publishing Company, 1975, pp. 101-125.

This page left blank

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov
Fred Oppel	6555	fjoppel@sandia.gov
Ben Cipiti	8845	bbcipit@sandia.gov
Lon Dawson	8851	ladawso@sandia.gov
Andrew Hahn	8851	ashahn@sandia.gov
Lee Maccarone	8851	lmaccar@sandia.gov
Benjamin Liu	8854	brliu@sandia.gov

Email—External

Name	Company Email Address	Company Name
Daniel Warner	daniel.warner@nuclear.energy.gov	DOE-NE

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.