# Empirical Risk Analysis Methodology for Adversarial Threats against Critical Infrastructure

David W. Wallace, A.M.ASCE[1]; Lloyd Foster[2]; Kris Schartau, A.M.ASCE[3]; David Lewin, A.M.ASCE[4]; and Conrad G. Keyes Jr., P.E., D.WRE, L.S., Dist.M.ASCE[5]

**Abstract:** The increase in foreign and domestic threats mandates a serious reevaluation of existing security methodologies, standards, and vulnerability assessments. A comprehensive defense strategy with quantitative and qualitative measurements is presented on how the water sector can optimize the application and placement of physical security countermeasures to improve resilience based on known parameters in a cost effective way. This study reviews the history and original intent of these methodologies that were adopted from the atomic and nuclear segments of the energy sector. These methodologies served as a starting point for the risk assessment documents that govern water sector security. The current American National Standards Institute (ANSI) risk models used by the water sector, based on design basis threat (DBT) and risk analysis and management for critical asset protection (RAMCAP), are rooted in the traditional risk formula of threat multiplied by vulnerability multiplied by consequence. This paper concludes that due to the inability to define who the adversary is, along with their objectives, motives, and capabilities, and the lack of statistically valid datasets or available intelligence of malevolent threats, the requirements listed in these methodologies are not achievable and will remain as unknowns in water/wastewater/stormwater systems. Therefore, the risk models used for mitigating adversarial threats have fundamental errors that should be replaced by an alternate risk model capable of measuring what can be known about facility resilience to malevolent attacks. By treating risk as a vector quantity consisting of known parameters, the probability of success of a given threat can be calculated using the mathematical analysis of defense strategy and countermeasures (MADSC) methodology. Once these parameters are established, the MADSC methodology can be used to determine the degree of difficulty in compromising existing countermeasures and provide guidance for physical security improvements and budgeting based on quantitative results. **DOI: 10.1061/JITSE4.ISENG-2291.** 



## Introduction

The water sector has experienced an inadvertent gap in physical security by using risk models not suited for the specific needs of community water systems based on design basis threat (DBT) and risk analysis and management for critical asset protection (RAMCAP). While these risk models can provide important aspects to consider, they are founded on the characterization of an adversary's objectives, motives, capabilities, and the probability of adversarial threats based on historical frequency, or available intelligence. The risk models do not adequately assess the security of a facility should an unexpected attack occur. Resilience in the face of any attack, expected or unexpected, is best determined by a facility's security and not an estimated low probability of attack. Additionally, security countermeasures can be mathematically evaluated to help optimize facility resiliency and cost-benefit analysis.

This study reviews the history and original intent of these methodologies that were adopted from the atomic and nuclear segments of the energy sector. These adjusted methodologies became the basis for risk assessment and are referenced in the following documents.

- ANSI/ASCE/EWRI 56-10 (ASCE 2010b), *Guidelines for the Physical Security of Water Utilities*, Standard.
- ANSI/ASCE/EWRI 57-10, *Guidelines for the Physical Security of Wastewater/Stormwater Utilities*, Standard.
- ANSI/AWWA J100-10(R13), *Risk and Resilience Management of Water Wastewater Systems*, Standard (recommended by the America's Water Infrastructure Act (2018) requirement for Risk and Resilience Assessments).
- US EPA 817-F-21-004, *Baseline Information on Malevolent Acts for Community Water Systems Version 2.0*.
- ANSI/AWWA G430-14(R20) (AWWA 2020), *Security Practices for Operation and Management*, Standard.
- ANSI/AWWA G440-17 (AWWA 2017), *Emergency Preparedness Practices*, Standard.

Three systemic issues significantly limit the success of risk models promoted for the water sector throughout the presented documents. The first is the requirement that the adversarial type and capability must be defined according to the DBT. The second is that the probability of a malevolent threat should be based on historical frequency, or available intelligence if historical data is unavailable, according to the RAMCAP methodology. Third, by utilizing an all-hazards approach, mitigating the risk due to natural hazards has been confused with mitigating the risk due to malevolent threats.

[1]Executive Director, Critical Security Research and Analysis Center, 1196 Diamond Cir, Ste. R, Lafayette, CO 80026 (corresponding author). ORCID: https://orcid.org/0000-0002-9869-4391. Email: david.wallace@csrac.us

[2]Chief Executive Officer, Foster Colley Enterprises, 2121 Windy Hill Rd., #913, Marietta, GA 30060. Email: lloyd.foster@fostercolley.com

[3]Defense Strategy Analyst, Surveillance One, 1196 Diamond Cir, Ste. Q, Lafayette, CO 80026. ORCID: https://orcid.org/0009-0009-0934-2852. Email: kris.schartau@survone.com

[4]Critical Infrastructure Regional Sales Manager, Echodyne, 12112 115th Ave. NE, Ste. A, Kirkland, WA 98034. Email: dlewin@echodyne.com

[5]Professor and Department Head Emeritus, New Mexico State Univ. (NMSU), 801 Raleigh Rd., Las Cruces, NM 88005. Email: ckeyes@nmsu.edu

Note. This manuscript was submitted on December 29, 2022; approved on July 10, 2023; published online on November 6, 2023. Discussion period open until April 6, 2024; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, © ASCE, ISSN 1076-0342.

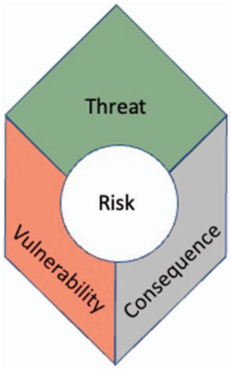


**Fig. 1.** (Color) Traditional risk components.

It is the opinion of the authors that these approaches have caused wide error bands and reduced confidence in the risk models, resulting in the low prioritization of physical security.

Traditionally, risk is comprised of three elements including threats, vulnerabilities, and consequences, as shown in Fig. 1. The traditional risk formula utilized by the water sector from the standards and guidelines previously listed consists of threat likelihood multiplied by vulnerability multiplied by consequence as shown in the following equation:

$$\text{Risk} = \text{Threat Likelihood} \times \text{Vulnerability} \times \text{Consequence} \quad (1)$$

The current approach to estimating threat likelihood for the traditional risk equation focuses on the ability to identify who the adversary is along with characterizing their objectives, motives, and capabilities. Statistically valid datasets or available intelligence on malevolent threats is deficient or arrives too late to prepare for an attack. Using best estimate or proxy analysis to determine threat likelihood can potentially lead to an exaggerated response or a false sense of security. Therefore, the risk models used for mitigating adversarial threats have fundamental flaws that must be replaced by an alternate risk model capable of measuring what can be determined about facility resilience.

By evaluating known criteria, the probability of success of a given threat ($P_{S|T}$) becomes a function of a set of triplets: scenario identified ($s_i$), difficulty identified ($d_i$), and consequence identified ($c_i$) (Wyss et al. 2011). The triplet can be treated as a vector quantity for the analysis process. Once the vector parameters are established, countermeasures can then be quantitatively and qualitatively measured, and the degree of difficulty to compromise can be empirically determined for cost-benefit analysis (Wallace and Foster 2021). This methodology removes the dependency of the previous requirements and improves guidance in vulnerability assessment recommendations and physical security budgeting.

### Historical Context

After 9/11 and the creation of the US Department of Homeland Security (DHS), the fight began to secure America's critical infrastructure against future attacks. If the United States could suffer an aerial attack from adversaries using our own airplanes against us, then what else could be possible? In the race to protect community water systems, DBT was quickly adopted from the US Nuclear Regulatory Commission's "1979 DBT Rule" as a security methodology. It was not recognized at the time that the DBT methodology contained criteria such as identifying the objectives, motives, and capabilities of an adversarial attack against the water sector were unattainable as described in the following section "Dilemma One: Characterization of an Adversary."

The RAMCAP methodology was first introduced to nuclear power plants (NPP) in 2005. In 2010, the ANSI/AWWA J100-10 document was published by the American Water Works Association (AWWA) and modified RAMCAP to create risk assessment methodology-water (RAM-W), adopted from Sandia Laboratories (AWWA 2013). The likelihood of malevolent threats based on historical frequency or available intelligence is discussed in the "Dilemma Two: Use of Frequentist Probability" section.

The RAMCAP methodology further identified threat characterization as all threats that included man-induced hazards or accidents, natural hazards, and dependency hazards (AWWA 2013). This is acceptable in general terms, but malevolent acts and random natural weather events should not be summarized into the same reference hazards as they are in the Summary of RAMCAP Reference Hazards in the ANSI/AWWA J100-10(R13) document (AWWA 2013). An all-hazards approach has treated natural disasters such as floods, tornadoes, or hurricanes alongside man-induced accidents and malevolent acts such as vandalism, criminal, and sabotage, and classified them together as reference hazards. Simply stated, natural hazards should be listed independently as different design-based events, not combined with malevolent threats. This is discussed in the "Dilemma Three: The Design Basis Event versus Design Basis Threat Confusion" section.

Subsequently, the ANSI/AWWA G430-14(R20), *Security Practices for Operation and Management*, Standard, the ANSI/AWWA G440-17, *Emergency Preparedness Practices*, Standard, and the EPA 817-F-21-004, *Baseline Information on Malevolent Acts for Community Water Systems Version 2.0*, documents were all built upon these foundations and followed suit.

### Dilemma One: Characterization of an Adversary

First, the DBT methodology assigns a security approach to defend against a hypothetical attack based on the objectives, motives, and capabilities of a potential assailant according to the adversarial classifications of Vandal, Criminal, Saboteur, or Insider (ASCE 2010a, b). The persistent problem has been knowing which countermeasures to select given the near impossibility of determining who the adversary might be and the assessment of their objectives, motives, and capabilities.

The process was meant to identify adversarial types, means, methods, motivations, and the capabilities of adversarial threats. This identification could also include various modes of attack with explosives (air, land, water) of various sizes (small, medium, large, and extra-large) and attacks not involving explosives (contamination, theft, and cyberattacks) (AWWA 2013). While it is important to discover what an adversary might scheme, it is not cost effective to overharden every defensive security layer for what could happen. However, the challenge remains for how key stakeholders can apply a deterministic approach to physical security within the water sector independent of adversarial characterization.

### Dilemma Two: Use of Frequentist Probability Theory

Second, the method of scoring threats, vulnerabilities, and consequences would seem reasonable, but this approach was to base threat likelihood on historical data. According to the RAMCAP model, risk is defined as (AWWA 2013)

$$\begin{aligned}\text{Risk} = {} & \text{Likelihood (Specific Attack)} \\ & \times \text{Vulnerability (Specific Attack)} \\ & \times \text{Consequence (of Attack)}\end{aligned} \quad (2)$$

The idea that threat likelihood could be calculated based on historical frequency was based on how natural hazards were evaluated. However, with malevolent threats, the actions of reconnaissance, planning, financial preparations, acquiring insider knowledge, and executing threat plans are all intentional steps, not random events such as natural disasters or accidents.

The study of historical data for determining probability is known as frequentist probability and has been defined in the DHS Risk Lexicon-2010 Edition that states in the annotation that (McNamara and Beers 2010)

1. Within the frequentist probability interpretation, precise estimation of new or rarely occurring events, such as the probability of a catastrophic terrorist attack, is generally not possible.
2. Frequentist probabilities generally do not incorporate "degree of belief" information, such as certain types of intelligence information.

A statistically valid set of data does not exist to make a frequentist approach to malevolent threats relevant as it does with natural hazards. This is illustrated in the US EPA document that assigns the value of $1 \times 10^{-6}$ as the annual default threat likelihood for multiple threat scenarios against water and wastewater facilities (US EPA 2021). If $1 \times 10^{-6}$ is used for the threat likelihood value in the risk equation, the resulting low score places physical security among the lowest priority of mitigations for threat hazards. Security practitioners have a difficult time justifying the cost of appropriate countermeasures with this kind of risk estimation.

The problem in this scenario is that if a particular region has never experienced extreme adversarial threat scenarios such as parachuting from helicopters, then the logical conclusion provided in vulnerability assessments was the recommendation of low monetary investment toward physical security countermeasures. This simply illustrates that there has not been enough data to prove that deadbolts or door locks should be installed and emphasizes that there is a logic problem with the frequentist methodology approach used for malevolent threats. It has been the experience of the authors that many individuals responsible for physical security have recognized that risk estimations like this did not seem accurate and deployed their own logical security choices.

The ANSI/AWWA J100-10(R13) does state that "intelligence information from law enforcement and intelligence agencies can be extremely useful in estimating threat likelihood, but the absence of such data should not deter the analyst from making the estimate based on the best available information" (AWWA 2013). Historical frequency (likelihood) and available intelligence are critical to the success of the RAMCAP model.

But who maintains the responsibility for available intelligence, guarantees its accuracy, or deploys the appropriate defense strategy in a timely manner? Should intelligence point to an attack, it might be too late to respond or prepare; therefore, implementing resilient solutions to start with is a better path to long-term security.

### *Dilemma Three: The Design Basis Event versus Design Basis Threat Confusion*

Third, the resiliency guidance documents list random weather events alongside adversarial threats such as assailants parachuting from helicopters to inflict damage on critical assets. The intent was to classify what events posed the greatest threat to the local provision of finished water in order to invest capital appropriately to meet the threat. But to the contrary, earthquakes in certain regions provided more of an existential threat than adversaries parachuting from helicopters, and funds were allocated for seismic mitigation rather than physical security based on the risk assessment.

Design basis events (DBE) are impacting events that serve as a basis of design for resilience and safety against natural disasters. DBEs are defined into subcategories of random natural disasters and sometimes even manmade accidents. Examples can be designs based on hurricanes, tornadoes, floods, earthquakes, fires, or other natural hazards (US NRC 2021). The application of design-based events to natural disasters works well because of the availability of known historical data. However, DBE models do not work well when applied to malevolent human threats because of the lack of known data, especially when applied to the specificity of who the adversaries are, such as vandals, criminals, saboteurs, insider threat, cyber insider/outsider threats, or terrorists, and their characteristics and capabilities.

Understanding how to appropriately classify design-based events is necessary when building a resilient design to counter potential catastrophic events. In water construction projects, design and engineering firms reference established natural disaster and weather-related statistics to help build water infrastructure capable of withstanding the design basis event they anticipate will happen in the future. For instance, if historical data convey that an impacting flood event occurs once every ten years, then there is a one in ten (10%) chance that a flood will occur at least once within a given decade. However, this methodology fails when applied to human behavior, apart from accidental, because it is not a random event. This is because the threat of an adversary involves:

- Intentional surveillance of the critical asset to be compromised.
- Planning for how to defeat protective systems.
- Education on how the system works to affect compromise.
- Financial capability to back software development or compromise tools as needed.
- Ability to deploy them all without being detected or arrested by authorities in the process.

A US interagency security committee, made up of 58 federal departments and agencies chaired by DHS, provides a regularly updated DBT report that describes "undesirable events" ranging from theft to active shooter that places humans at the center of a DBT (Durkovich 2016). All these processes are far from random and do not fit in randomization calculus. Therefore, adversarial threat scenarios should not be listed alongside random natural phenomena.

To further complicate this process, a decision made by an adversary one day might be completely different than a decision made on the next day based on mood, external influence, or simply added information being brought to the scenario (Bayes' theorem). Each juncture in the decision tree has an unlimited array of possible next decisions that are now a part of the equation. Mathematically and practically speaking, it is impossible to predict or try to solve all the decisions that an adversary might make. For example, if a water utility were to build a bigger and more difficult-to-compromise perimeter wall than the traditional fence fabric, an adversary would have to consider many options to breach the wall. The adversary could choose to scale the wall, go around the wall, tunnel underneath the wall, use explosives to blow a hole in the wall, drive a hardened vehicle through the wall, acquire a taller ladder, make an aerial entrance to bypass the wall, or just be deterred and give up altogether and choose another target elsewhere. This idea is further illustrated in *Risk-Based Cost-Benefit Analysis for Security Assessment Problems*: "However, for high-security facilities, security risk is much harder to quantify than safety risk since the probability of attack is highly uncertain and depends strongly on unquantifiable psychological factors such as deterrence and adversary goal intensity" (Wyss et al. 2009).

### Significant Developments to the Risk Equation

In 1981, the document *On the Quantitative Definition of Risk*, written by Kaplan and Garrick, proposed using "triplets" as a vector quantity to describe risk as a set of probabilities, scenarios, and consequences as shown in the following equation (Kaplan and Garrick 1981):

$$\text{Risk} = \{s_i, p_i, c_i\} \tag{3}$$

The basis of risk analysis involved answering three questions that made up each element of the triplet.

1. What can happen? [The identified scenario ($s_i$)].
2. How likely is it that [it] will happen? [The probability of the scenario ($p_i$)].
3. If it does happen, what are the consequences? [The consequence of the scenario ($c_i$)] (Kaplan and Garrick 1981).

Here, the probability of the scenario could be applied to natural disasters or even manmade accidents, but the estimation of adversarial threats still plagued the problem since determining the probability of the scenario was based on historical frequency, which could not be determined.

In 2010, a document called *A Risk Informed Method for Enterprise Security (RIMES)* was introduced by Sandia National Laboratories where Wyss et al. (2011) resurfaced the Kaplan and Garrick triplets with a modification that leveraged the approach by replacing the probability of the scenario ($p_i$) with the degree of difficulty ($d_i$). The degree of difficulty refers to how hard it would be to successfully accomplish the scenario against the target under consideration. This means analyzing how many countermeasures would have to be compromised for the adversary to reach the critical asset (Wyss et al. 2011). By measuring what can be known instead of attempting to utilize historical frequency data that are not available for malevolent threats, this dramatically improved the risk equation.

If an adversary on foot were to choose a path of jumping over a fence to initiate the compromise of a targeted asset (scenario), with the intent to cause catastrophic failure of a water treatment plant (consequence), then the countermeasures required to be defeated would represent the degree of difficulty to complete the task. If the triplets for security risk $s_i$, $d_i$, $c_i$ are known, then they become a function of the conditional probability of success of a given threat, $P_{S|T}$ as shown in the following equation:

$$P_{S|T} = f(s_i, d_i, c_i) \tag{4}$$

By holding the conditional probability of success of a given threat ($P_{S|T}$) constant at a value indicative of adversary success, the degree of difficulty ($d_i$) over a broad range of scenarios ($s_i$) is determined, including the threshold threat characteristics required for an adversary to be successful in the scenario ($s_i$) (Wyss et al. 2011). This measurement cannot stem from evaluating a dataset of adversarial capability as the data do not exist. And even if the data did, there are an infinite number of variables that could be introduced into the equation that would significantly complicate the formula. The degree of difficulty must be measured across the countermeasures of the defense layers to determine what all would have to fail for a complete compromise of the facility. The DHS Risk Lexicon-2010 defines vulnerability as the "qualitative or quantitative expression of the level to which an entity, asset, system, network, or geographic area is susceptible to harm when it experiences a hazard" (McNamara and Beers 2010). Referencing the annotation associated with the definition in the DHS risk lexicon, the degree of vulnerability of the facility is the conditional probability of success of a given threat as shown in the following equation:

$$\text{Vulnerability} = P_{S|T} \tag{5}$$

### Cumulative Defense Strategy

With a few small adjustments, nonoffensive defense strategies can be made more effective. The defense in depth methodology uses diverse protective measures along each potential adversarial path (Sandia National Laboratories 2016), but this can be further enhanced by requiring the increase of the quality and quantity of countermeasures throughout the scale of each defense layer. The authors have termed this cumulative defense strategy. Incrementally adding quality and quantity of countermeasures throughout all the defense layers increases the required resources for an adversary to succeed and consequently addresses the variance in adversarial capability. This incremental increase establishes the minimum difficulty threshold (MDT) level required at each step. Once this process is complete, it enables a consistent approach on which mathematical analysis can be performed to determine the probability of success of a given threat and a cost-benefit analysis based on the existing countermeasures (Wallace and Foster 2021).

To better understand cumulative defense strategy, a basic review of defense strategy is important for classifying attack vectors and the counter-defense layers against them. Additionally, CDS includes the ordinal steps an adversary must sequentially cross to reach the critical asset and the defense layer elements that oppose them and the countermeasures with specific characteristics to help defend against the attack.

### Attack Vectors

The attack vectors define how an adversary might attack the critical asset. Identifying the attack vectors does not attempt to characterize the adversary; it simply states the type of "vehicle" an adversary might use to carry out the attack. A variety of attack vectors might include:

- Adversary on foot (AOF)
- Vehicle ramming breach (VRB)
- Vehicle-borne improvised explosive device (VBIED)
- Adversarial use of maritime vehicles
- Unmanned (uncrewed) aircraft systems (UAS)–surveillance, data sniffing, weaponization

Manned aircraft assaults on facilities have been omitted. Although detection capability is possible for aircraft, there are no reasonable countermeasures available to protect against this type of attack. The water sector cannot be expected to protect against threats that are of a national security nature and the responsibility of the US Federal Government to defend against.

### Defense Layers and Defense Layer Elements

Defense layers comprise defense layer elements and countermeasures to protect the critical asset. They are in alignment with the ordinal steps necessary to breach the target. The defense layer elements are the physical locations at the facility in combination with strategic countermeasures designed to oppose specific attack vectors. Table 1 includes an example set of defense layer elements for some sample water facilities.

### Characteristics of Countermeasures

Understanding the characteristics of countermeasures will help guide the selection of countermeasures to be used within each of the defense layer elements. Countermeasures embody one or more of the following characteristics:

- Deter—Used to dissuade the occurrence of low-level attacks. No considerable effectiveness.

**Table 1.** Defense layer elements for small, medium, and large facilities

| Defense layer element | Small facility (pump stations/tanks) | Medium facility (>50 k population served) | Large facility (>100 k population served) |
|---|---|---|---|
| 1 | Perimeter | Perimeter | Perimeter |
| 2 | Detection line | Detection line | Detection line |
| 3 | Observation zone | Observation zone | Observation zone |
| 4 | Critical asset ingress | Facility exterior ingress | Facility exterior ingress |
| 5 | — | Critical asset ingress | Facility interior |
| 6 | — | — | Critical asset ingress |

- Detect—Used for detecting an intrusion via physical alarm trigger, virtual tripwire, or analytics.
- Observe—Used for maintaining visual awareness through optics, thermal, or radar capability.
- Delay—Used for extending the length of time required for an adversary to progress.
- Deny—Used to create controlled access and authorized user levels.
- Respond—Use of sequential steps to respond to an adversary or prepare for potential compromise.

### *Mathematical Analysis of Defense Strategy and Countermeasures*

The key to unlocking the application of this risk methodology was discovered in 2021 by Lloyd Foster and David Wallace by mathematically calculating the degree of difficulty to achieve the probability of success of a given threat for a scenario identified ($s_i$), difficulty identified ($d_i$), and consequence identified ($c_i$). The new math model analyzes the current countermeasures across the defense layer elements and determines the optimized placements mathematically and objectively for improvements. The process of evaluation is called the mathematical analysis of defense strategy and countermeasures (MADSC), and it requires the sequential increase of quantitative and qualitative countermeasures (Wallace and Foster 2021). Defense layer elements and countermeasures are strategically placed to address specific attack vectors. Path analysis of each attack vector is then evaluated for the number of ordinal steps (defense layer elements) and the subset of countermeasures within each of these ordinal steps.

For instance, an adversary on foot might have to breach six ordinal steps to reach a critical asset, and each step consists of multiple countermeasures that collectively bolster each ordinal step. The steps are ordinal in nature because they must be sequentially crossed to reach the critical asset. At each ordinal step, the quantity of countermeasures must increase one number greater than its parent ordinal step to maintain a growing MDT, which in turn increases the defensive quality of each ordinal step. For example, ordinal step number one must contain at least two countermeasures, and ordinal step number two must contain three countermeasures, and so on. As referenced, this additive nature is known as the cumulative defense strategy (Wallace and Foster 2021).

The MADSC methodology is used to calculate the joint probability of compromise through the coupling of two different math models, which in Latin is called a "copula." Copulas were invented in 1959 by Abe Sklar, and about a dozen formulas have been invented since then to solve unique needs. This version is called the Foster-Wallace formula. The Foster-Wallace formula is unique in that it couples the probability density function (PDF) of a geometric distribution across the ordinal steps that are required to be defeated by an adversary with the cumulative distribution function (CDF) of a gamma distribution across the total defensive countermeasures within the subsets of the ordinal steps (Wallace and Foster 2021). The PDF of a geometric distribution represents the probability of an assailant reaching a specified ordinal step within the defense system, while the CDF of a gamma distribution represents the probability of an assailant bringing down the whole defense system by successfully breaching a given countermeasure. Coupling the distributions together provides the ability to generate the joint probability of compromise or the degree of difficulty to compromise the countermeasures at each ordinal step. The mathematical expression of the Foster-Wallace formula appears in the following equation (Wallace and Foster 2021):

$$f[x,\alpha,\beta,\rho,\gamma] := \left(\int_0^x \frac{e^{-\frac{z}{\beta}} z^{-1+\alpha}\beta^{-\alpha}}{\text{Gamma}(\alpha)}\,dz\right)^{((1-\rho)^\gamma\rho)^{(1-\rho)^\gamma\rho}} \tag{6}$$

where:

- $f$ = the function of the following parameters resulting in the joint probability of compromise.
- $\chi$ = the accumulated score assigned to a countermeasure (e.g., 1, 3, 6, and 10).
- $\alpha$ = the shape parameter of the gamma distribution used to model countermeasures.
- $\beta$ = the scale parameter of the gamma distribution used to model countermeasures.
- $\rho$ = the parameter of a geometric distribution used to model ordinal steps.
- $\gamma$ = an ordinal step (e.g., 1, 2, 3, and 4).

To understand the value of "$\chi$," the accumulated score is measured within the gamma distribution. Consider an asset that has six defense layer elements also known as ordinal steps. The accumulated ordinal score becomes the sum of all the ordinal steps that is $(1+2+3+4+5+6)=21$. Based on CDS, starting at the outermost defense layer, ordinal step N = 1, there must be a minimum of $N+1$ countermeasures at each ordinal step. Table 2 shows the correlation between each ordinal step, accumulated ordinal value, and countermeasure number for scenarios with up to six ordinal steps. It is not necessary to understand the mathematical complexities of the formula to understand the results of the MADSC methodology. The results are clear in revealing locations that are well protected and exposing areas that need specific improvements. In either case, an actionable optimization of countermeasures is provided.

The output of the Foster-Wallace formula shown in Fig. 2 graphically depicts the joint probability with respect to each defense layer element and its corresponding countermeasures using a 3D plot. It is important to note the color change within the plot. Blue correlates to the lowest degree of difficulty, while red correlates to the highest degree of difficulty. The plot is best interpreted starting at the front lower left corner and working back to the upper right corner. Starting at the perimeter, the degree of difficulty penetrating the perimeter fence and crossing the observation zone is comparatively low, as shown by the blue color on the plot. Once the facility exterior is reached, the rate of rise in the degree of difficulty begins to grow rapidly because the quantity and quality of countermeasures have increased and are more difficult to defeat. At defense layer element 4, the countermeasures embody more delay and deny characteristics, forcing the adversary to expend more resources to make it to the next defense layer element. At each new defense layer element, the quantity and quality of countermeasures are increasing, thereby reducing the probability the adversary will defeat all the countermeasures required to breach the critical asset.

The math model generates different degrees of difficulty values depending on the total number of defense layer elements for the facility. Referring to the example facilities in Table 1, each facility

**Table 2.** Correlation between ordinal steps, and countermeasures

| Ordinal steps | Accumulated ordinal values | Countermeasure number $(N+1)$ |
|---|---|---|
| 1 | 1 | 1 |
| | | 2 |
| 2 | 3 | 3 |
| | | 4 |
| | | 5 |
| 3 | 6 | 6 |
| | | 7 |
| | | 8 |
| | | 9 |
| 4 | 10 | 10 |
| | | 11 |
| | | 12 |
| | | 13 |
| | | 14 |
| 5 | 15 | 15 |
| | | 16 |
| | | 17 |
| | | 18 |
| | | 19 |
| | | 20 |
| 6 | 21 | 21 |
| | | 22 |
| | | 23 |
| | | 24 |
| | | 25 |
| | | 26 |
| | | 27 |

has a different number of defense layer elements depending on the size and type of critical asset requiring protection. Typically speaking, larger facilities with less redundancy pose a greater risk to the community if they were compromised, so strategically applying more countermeasures to defend the critical asset should reduce the probability of success of a given threat. Fig. 3 shows the degree of difficulty generated by the math model for facilities requiring four, five, and six defense layer elements. At each ordinal step, the countermeasure number corresponds to a different degree of difficulty depending on the defense layer elements considered for the facility.

Table 3 shows a summary output of the Foster-Wallace formula based on an example facility with six defense layer elements. Plotting the degree of difficulty to compromise against the countermeasure number from Table 3 yields the 2D graph shown in Fig. 4. The countermeasures called out in Fig. 4 are suggested but are in no way prescriptive. They do, however, illustrate the importance of increasing the quantity and quality of physical security countermeasures throughout the facility. Defense strategy analysts can begin to determine security vulnerabilities within the facility by calculating the degree of difficulty based on a known set of countermeasures, defense layer elements, and attack vectors. Fig. 4 can be used as a working model for larger facilities, but minor calibrations might need to be performed per site to adjust for smaller facilities only requiring four or five defense layer elements as depicted in Fig. 3. This model has been used numerous times by the authors, and the results show a high-resolution visibility into the effectiveness of countermeasures and the capability level of a facility's defense strategy.

Intuitively, the shape of the curve in Fig. 4 makes sense. The first defense layer element, ordinal step one, typically consists of a perimeter fence and "No Trespass" signs as countermeasures. Defeating a fence and signs represents a low degree of difficulty as shown in the plot. Ordinal step two, detection line, exemplifies the need to detect the presence of an adversary. Countermeasures throughout ordinal step two typically embody strong deter and detect characteristics but offer little stopping power to a determined attacker. The observation zone, ordinal step three, employs countermeasures that attempt to put eyes on the target to track their movement. Starting at ordinal step four, the plot begins to rise rapidly and is reflected by the facility's exterior ingress. One would expect a greater degree of difficulty trying to gain access to a secure facility. The quantity and quality of countermeasures have increased, thereby increasing the degree of difficulty to proceed further into the facility. If the adversary breaches the facility exterior, then they have reached ordinal Step 5 known as the facility interior. Here, the rate of rise is mostly consistent representing additional countermeasures with deter, detect, observe, delay, deny, and respond characteristics. If the

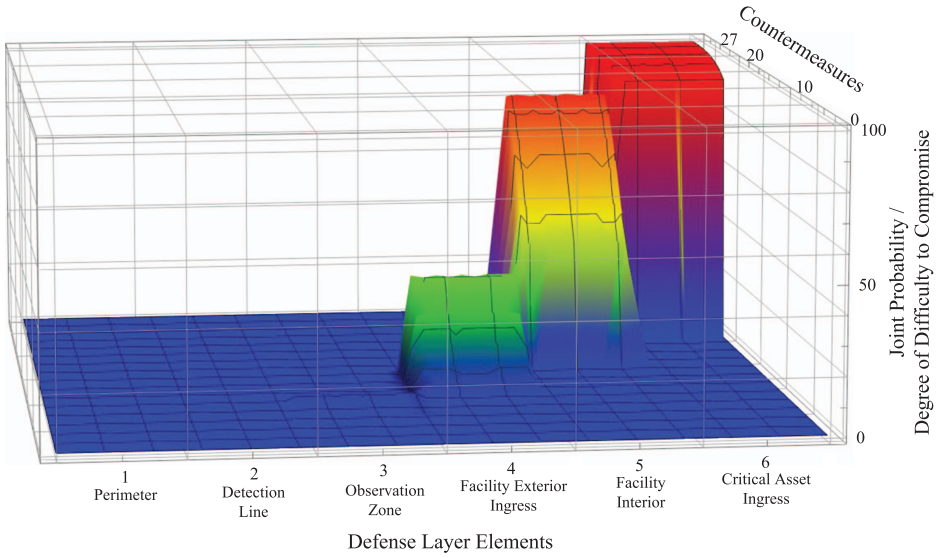


**Fig. 2.** (Color) MADSC 3D math model for six defense layer elements.

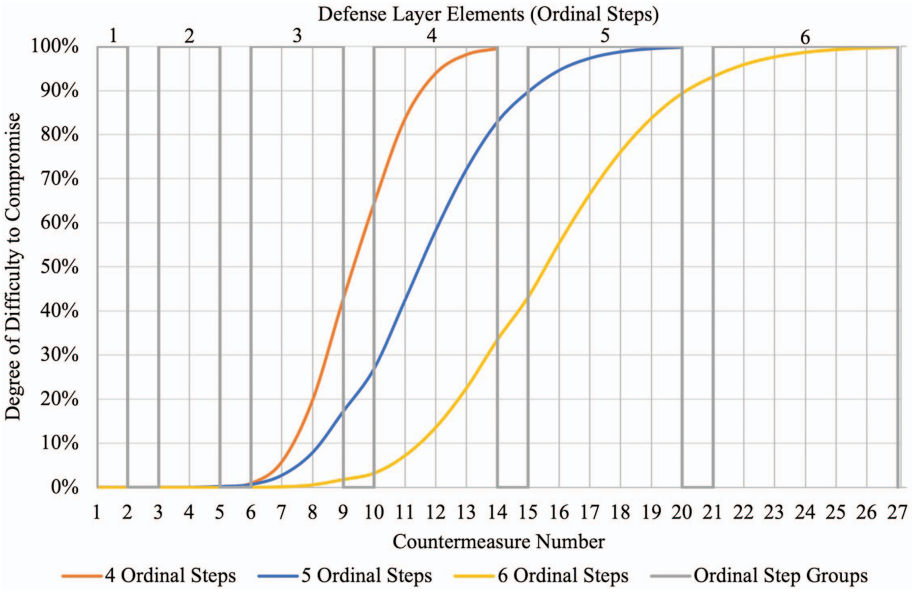


**Fig. 3.** (Color) Foster-Wallace formula output for the degree of difficulty of four, five, and six defense layer elements.

**Table 3.** Output of the Foster-Wallace Formula based on a facility with six defense layer elements

| Ordinal step | Geometric distribution PDF | Countermeasure number | Gamma distribution CDF | Joint probability | Defense layer elements up to breach |
|---|---|---|---|---|---|
| 1 | 0.243869 | 1 | $6.96715 \times 10^{-21}$ | $2.06463 \times 10^{-15}$ | Perimeter |
| 1 | 0.243869 | 2 | $2.72687 \times 10^{-14}$ | $1.31168 \times 10^{-10}$ | |
| 2 | 0.191788 | 3 | $1.08495 \times 10^{-10}$ | $8.64423 \times 10^{-8}$ | Detection line |
| 2 | 0.191788 | 4 | $2.57682 \times 10^{-8}$ | $4.17535 \times 10^{-6}$ | |
| 2 | 0.191788 | 5 | $1.30976 \times 10^{-6}$ | $6.76153 \times 10^{-5}$ | |
| 3 | 0.102839 | 6 | 0.0000251711 | 0.023% | Observation zone |
| 3 | 0.102839 | 7 | 0.000247888 | 0.140% | |
| 3 | 0.102839 | 8 | 0.00150058 | 0.582% | |
| 3 | 0.102839 | 9 | 0.0062819 | 1.809% | |
| 4 | 0.0433666 | 10 | 0.0197211 | 3.250% | Facility exterior ingress |
| 4 | 0.0433666 | 11 | 0.0492183 | 7.220% | |
| 4 | 0.0433666 | 12 | 0.102017 | 13.640% | |
| 4 | 0.0433666 | 13 | 0.181679 | 22.571% | |
| 4 | 0.0433666 | 14 | 0.285624 | 33.499% | |
| 5 | 0.0182876 | 15 | 0.405327 | 43.200% | Facility interior |
| 5 | 0.0182876 | 16 | 0.529002 | 55.331% | |
| 5 | 0.0182876 | 17 | 0.645183 | 66.545% | |
| 5 | 0.0182876 | 18 | 0.74552 | 76.113% | |
| 5 | 0.0182876 | 19 | 0.825926 | 83.715% | |
| 5 | 0.0182876 | 20 | 0.886181 | 89.377% | |
| 6 | 0.0077118 | 21 | 0.928691 | 93.122% | Critical asset ingress |
| 6 | 0.0077118 | 22 | 0.957084 | 95.863% | |
| 6 | 0.0077118 | 23 | 0.97513 | 97.604% | |
| 6 | 0.0077118 | 24 | 0.986089 | 98.660% | |
| 6 | 0.0077118 | 25 | 0.992473 | 99.275% | |
| 6 | 0.0077118 | 26 | 0.996053 | 99.620% | |
| 6 | 0.0077118 | 27 | 0.997989 | 99.806% | |

adversary has successfully reached the critical asset ingress, ordinal Step 6, this would be the last step before a total compromise of the critical asset. Installing countermeasures before this step that embody strong delay and deny characteristics becomes critical. At this step, the rate of rise on the plot decreases significantly as a full breach approaches.

It is important here to note that because certain defensive countermeasures score very low in their effective difficulty, this does not mean that they are unimportant or should not be used. A "No Trespass" sign is effectively useless in physically stopping someone from entering a site, so it scores exceptionally low on this chart. However, "No Trespass" signs are an important countermeasure to provide a warning for those who would consider trespassing and additionally serve a legal purpose that a lawful notice was served on the premise.

Similarly, security technology such as surveillance cameras, thermal cameras, and ground radar devices can garner an unwarranted perception of security. Although they are a necessary part of

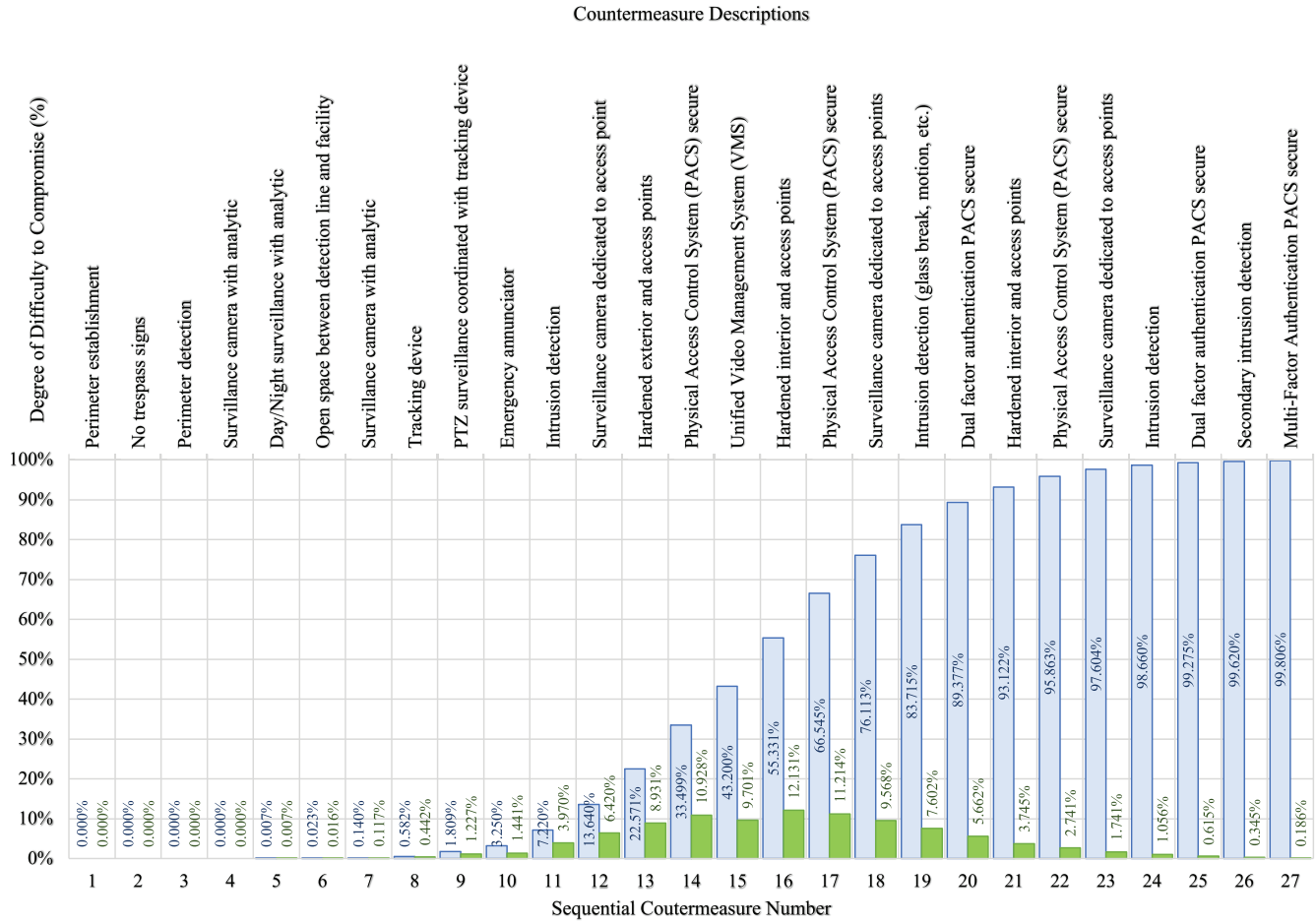


**Fig. 4.** (Color) Example set of possible countermeasures placed across six defense layer elements.

the overall defense strategy, by themselves they do not provide any real stopping power toward a malevolent actor. In the characteristics of countermeasures, surveillance cameras can provide the ability to deter, detect, and observe but not delay, deny, or respond.

Fig. 4 can be used to determine the resiliency of a site by comparing the existing countermeasures with what should be in place according to the chart. As an example, if all countermeasures were in place up through countermeasure #24, then the site would receive a 98.66% degree of difficulty to compromise rating (Fig. 4). Or otherwise stated, an adversary would have to express 98.66% of the overall difficulty required to breach countermeasure #24. However, if the access control system was still operating at 125 MHz, which is an antiquated version of access control that can now be easily compromised with duplicated cards (like keys), and then, credit for the access control features should not be granted. In this case, it would be appropriate to subtract from the overall joint probability score the access control scores for all ordinal steps that do not support the secure access control standards of 13.56 MHz. Since three ordinal steps require this, their total is $24.80 = (10.93 + 12.13 + 1.74)$ must be subtracted from 98.66 to equal a 73.86 degree of difficulty score.

### *Cost-Benefit Analysis Approach*

The cumulative defense strategy method also enables a consistent approach by which mathematical analysis can be performed to determine the probability of success of a given threat and the evaluation of cost-benefit analysis. Defense strategy analysts can quickly identify security deficiencies in facilities using the MADSC methodology. The model shows where countermeasures have the greatest impact to potentially defeat an adversary. According to the model, installing fiber fence technology at the perimeter provides detection of an intrusion but has little effect on the overall joint probability. Avoiding this cost in exchange for less expensive surveillance cameras with built-in detection analytics that are positioned to deter, detect, and observe the adversary is in many cases an acceptable choice. Every facility is physically unique and will require explicit analysis to ensure the critical asset is protected in the best way possible. Some facilities are in a dense urban area with unassociated buildings and public streets on all sides. The exterior of the facility must inherit the characteristics of the perimeter, detection line, and observation zone to create a combined defense layer. Countermeasure selection in this situation begins to focus on what provides the best deter, detect, observe, delay, deny, and respond characteristics. Reducing the number of ingress points to the facility will reduce the total number of countermeasures required to protect the facility.

## Conclusion

Unfortunately, most risk assessments currently written for the water sector are based on risk models that are likely to register flawed results. Attempting to identify an adversary's characteristics or the probability of adversarial threats based on historical frequency or available intelligence has often resulted in underinvestment of

physical security countermeasures. In some cases, those who were aware of the problem may have overcompensated by deploying expensive countermeasures resulting in a false sense of security.

The objective is to strike a responsible balance between the extremes of normalcy bias, where many have embraced the belief that nothing ever really happens so nothing ever will, and the overreaction of recklessly investing in countermeasures without the guidance of an empirical methodology. By using the MADSC methodology, countermeasures can be quantitatively and qualitatively measured, and the degree of difficulty can be empirically determined and optimized. Additionally, the percentage remaining can be used to represent the remaining difficulty required for a complete compromise of the countermeasures and potentially the critical asset. Therefore, the placement for the best use of investment becomes clear for guarding against compromise and achieving the highest levels of intruder delay to allow for an adequate response time.

This new methodology contributes to the overall body of knowledge for the profession by providing an effective way of screening the various capabilities of potential adversaries and provides fiscally sound physical security practices for cost-benefit analysis and budgeting. Determining the risk facilities might face due to a malevolent attack using the MADSC methodology helps clarify how committed operators are to ensuring the safety of their water systems. Compromising with risk acceptance should only be a last resort unless other compelling options are available. The implementation of the MADSC methodology in the water sector could help lead the way in physically securing other DHS critical infrastructure sectors.

## Data Availability Statement

Some data, models, or code that support the findings of this study, such as Mathematica output PDF, are available from the corresponding author upon reasonable request.

## Acknowledgments

Special acknowledgment and thanks to Gregory D. Wyss, Ph.D., from Sandia National Laboratories who with his team has helped pioneer some of the concepts mentioned in this study. His generous time spent with the CSRAC team is much appreciated. Additionally, special acknowledgment and thanks to computational mathematician Lloyd Foster for his applied work with cumulative distribution function and probability density function in this proposed methodology stemming from his 20-year expertise in risk analysis and actuarial science.

## Disclaimer

The authors do not, and will not, guarantee that the implementation of risk mitigation techniques by the reader based on the information included in this document will eliminate a current or future risk/threat. The information in this document is provided for general informational purposes only in accordance with the customary professional standards within the industry.

The information in this document is predicated solely upon security issues known to the authors at the time of writing. The reader recognizes and understands that this document cannot and does not address all potential threats or risks, or their potential impact, as security, safety, emergency management, and crime prevention/reduction strategies are dynamic processes. As facility conditions are modified and/or expanded, and since threats often and commonly continuously change, it is imperative that the reader routinely review, update, and change their security process management, technology, policies, and procedures as necessary to reflect changes in the environment, the requirements of their business, and the expectations of members of the community.

## References

ASCE. 2010a. *Guidelines for the physical security of wastewater/stormwater utilities*. ANSI/ASCE/EWRI 57-10. Reston, VA: ASCE.

ASCE. 2010b. *Guidelines for the physical security of water utilities*. ANSI/ASCE/EWRI 56-10. Reston, VA: ASCE.

AWWA (American Water Works Association). 2013. *Risk and resilience management of water and wastewater systems*. AWWA J100-10(R13). Denver: AWWA.

AWWA (American Water Works Association). 2017. *Emergency preparedness practices*. ANSI/AWWA G440-17. Denver: AWWA.

AWWA (American Water Works Association). 2020. *Security practices for operation and management*. ANSI/AWWA G430-14(R20). Denver: AWWA.

Durkovich, C. 2016. *The risk management process for federal facilities: An interagency security committee standard*. 2nd ed. Washington, DC: US Department of Homeland Security.

Kaplan, S., and J. B. Garrick. 1981. "On the quantitative definition of risk." *J. Risk Anal.* 1 (1): 11–27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x.

McNamara, P. A., and R. Beers. 2010. *DHS Risk Lexicon 2010 edition*, 23–25. Washington, DC: US Department of Homeland Security.

Sandia National Laboratories. 2016. "International training course on the physical protection of nuclear facilities and materials." In *Introduction to the design of physical protection systems*. Albuquerque, NM: Sandia National Laboratories.

US EPA. 2021. *Baseline information on malevolent acts for community water systems version 2.0*. US EPA 817-F-21-004. Washington, DC: US EPA.

US NRC (United States Nuclear Regulatory Commission). 2021. "Enclosure 1—Background design-basis events, design-basis information, and external events." Accessed December 5, 2022. https://www.nrc.gov/docs/ML1432/ML14328A170.pdf.

Wallace, D. W., and L. Foster. 2021. *A logical basis for cumulative defense strategy and the mathematical analysis of defense strategy & countermeasures (MADSC)*. Lafayette, CO: Critical Security Research & Analysis Center.

Wyss, G. D., J. F. Clem, J. L. Darby, K. Dunphy-Guzman, J. P. Hinton, and K. W. Mitchiner. 2011. *A risk informed method for enterprise security*, 1–4. Albuquerque, NM: Sandia National Laboratories.

Wyss, G. D., J. Darby, C. Silva, and A. Water. 2009. *Risk-based cost-benefit analysis for security assessment problems*. SAND2009-3568C. Albuquerque, NM: Sandia National Laboratories.