

Advanced Reactor Safeguards & Security

Top Event Prevention Analysis Using a Concept of Margin

Prepared for US Department of Energy

R. Youngblood¹ , D.P. Blanchard² , M. Diaconeasa³

1 Idaho National Laboratory ²Applied reliability Engineering Inc. ³North Carolina State University

> **September 2024 INL/RPT-24-80536**

INL/RPT-24-80536

Top Event Prevention Analysis Using a Concept of Margin

R. Youngblood D. P. Blanchard M. Diaconeasa

April 2024

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Error! Reference source not found.

Top Event Prevention Analysis Using a Concept of Margin

[R. Youngblood](#page-1-0) D. P. Blanchard M. Diaconeasa

September 2024

Idaho National Laboratory Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the U.S. Department of Energy Office of Nuclear Energy Under DOE Idaho Operations Office Contract DE-AC07-05ID14517

Page intentionally left blank

Executive Summary

The purpose of this project is to advance the state of practice of managing cybersecurity risk efficiently at nuclear power plants. With the passage of time, generic cybersecurity risk management processes continue to proliferate, but generic risk management processes are not optimal for nuclear power plants. It has been recognized historically that nuclear plants call for very careful analysis supporting risk management decisions, and great care in the imposition of costs on nuclear plants. Accordingly, since the Three Mile Island accident, nuclear plant risk management practice has evolved significantly to become much more scenario-based, which has enabled a much-improved focus on safety with simultaneous cost savings. Safety thinking has evolved beyond reasoning based on selected postulated challenges to investing significantly in understanding the full set of possible scenarios, and quantifying their likelihoods and consequences. This understanding enables focus on risk-significant scenarios. The term of art for this modern approach is "risk-informed."

However, although methods used to analyze *safety* have improved dramatically, the methods used in safety do not automatically do a good job in security. Our risk analyses offer useful insight into what systems we need to protect from cyberattack, but adversarial attacks pose fundamental challenges to the methods commonly used in safety analysis. Management of safety issues is relatively straightforward if we know what to expect, and for current-generation plants, we do; but in security, that sort of knowledge is not necessarily to be had even for current-generation plants, much less advanced plants. In particular, estimating "attack likelihood" is controversial at best. Given a postulated cyberattack, we can model possible resulting scenarios, but we are still learning what specific attacks to postulate in plants containing thousands of components, and what sorts of new failure modes – ones that we would not expect to see occurring randomly - might be causable by adversaries. The purpose of the present task is to advance the state of practice in modeling cyberattacks, and reasoning about what equipment to protect, and how much of it to protect, against cyberattacks.

The purpose of this task was:

… to bring together several ideas that are not usually examined together: the idea that many decisions should ideally consider multiple attributes (multiple performance figures of merit, or FOMs) rather than just one, which calls for modeling those figures of merit in mutually consistent ways; application of Top Event Prevention Analysis (TEPA) to look for equipment treatments that simultaneously do a good job on all figures of merit analyzed; and the idea of reasoning about system dependability without over-relying on traditional probability concepts that are problematic in a security context.

Given the well-known difficulties of characterizing "attack likelihood," a "margin" concept would be useful in cyber risk management, if we could formulate an applicable notion of it.

We still do not have an operational concept of margin against cyberattack analogous to "attack difficulty," but we have accomplished the following:

- collected useful background information about the concept of margin, including potentially useful elicitation processes that could be valuable in planning the tabletop exercises intended to be done in FY 25 to learn how to assess a margin/difficulty attribute for specific attacks;
- in the course of reviewing "margin," we identified a potentially useful path forward for cyber risk management through application of "decision-making under deep uncertainty" (DMDU). This family of methodologies has certain high-level compatibilities with Top Event Prevention Analysis (TEPA) and with Risk-Informed Management of Enterprise Security (RIMES);
- advanced the computational technology of applying TEPA efficiently in multi-attribute analyses.

Tabletop exercises are being planned for FY 25 to advance the state of practice in assessing difficulty of specific cyberattacks. The present developments will directly support those tabletops.

Page intentionally left blank

CONTENTS

FIGURES

TABLES

ACRONYMS

Page intentionally left blank

1. Introduction

1.1 Risk-Informed Assurance

If a particular scenario leads to adverse consequences, we need assurance^{[1](#page-12-2)} that that scenario will not occur; and if the adverse consequences are *very* severe, we need *very* high assurance that that scenario will not occur. In safety, this means that we have to prevent the basic events comprised in the subject scenario, applying well-understood engineering principles; in cyber, this means making scenarios difficult for adversaries to cause. We have partial knowledge about how to do that, but we have not yet settled on a useful metric of difficulty in cyber.

The present discussion is scenario-based in the sense that it presupposes the sort of analysis that first develops, and then reasons from, a notionally^{[2](#page-12-3)} complete scenario set, such as the scenario set that one develops in a Probabilistic Risk Assessment (PRA). This is to be contrasted with the safety reasoning that operated in the nuclear domain before PRA: the idea that we could postulate a few extreme events (such as "large loss-of-coolant accident coincident with loss of offsite power and the limiting single active failure") and conclude that a plant was safe if it could satisfy performance criteria conditional on those severe safety challenges. The shortcomings of that paradigm have steadily become clearer in the years since WASH-1400 (1975). [1] As of the mid-1990's, it became NRC policy [2] to make use of PRA in regulatory decision-making as follows:

The use of PRA technology should be increased in all regulatory matters to the extent supported by the state-of-the-art [*sic*] in PRA methods and data and in a manner that complements the NRC's deterministic approach and supports the NRC's traditional defense-in-depth philosophy.

The starting point for the present discussion is the observation that ["](#page-12-4)risk-informing"³ the safety aspects of risk management for nuclear plants has been highly beneficial, and the realization of analogous benefits in the security domain would be beneficial if we could achieve them.

By "risk-informing," we mean to invoke not just the scenario-set idea, but also the point that quantification of scenario likelihoods and consequences has supported decision-making in ways that were not widely anticipated^{[4](#page-12-5)} before PRA began to be undertaken in earnest. But doing this for security has been seen as problematic [5]:

¹ For present purposes, the concept of "assurance" is to be understood as "reasons to believe a particular claim." For example, an "assurance case" provides evidence and argument to someone such as a regulator who needs to believe particular safety claims about a proposed system in order to grant a license to the applicant, or to an investor who needs to believe availability claims about a particular system to justify revenue expectations.

² The word "notionally" appears for two reasons: first, because PRA's completeness issue is recognized. Even if we cannot absolutely prove completeness of a PRA, experience shows that we are better off reasoning with a PRA, and supplementing the PRA technical basis through consideration of margin and defense in depth, than we are reasoning without PRA. That said, it is not clear that safety-oriented hazard analysis techniques have adequately evolved to address cyberattack.

³ NRC Chairman Jackson [3] defined "risk-informed" as follows: A "risk-informed" approach means that, in the decision-making process, quantitatively derived risk information is considered along with other factors such as the need for defense-in-depth, good engineering practice, and operating experience. Risk information does not become the sole basis for a decision (that is, the decision is not "risk-based"), but rather provides a systematic way of identifying and comparing what is important and where uncertainties exist.

⁴ An exception was the early work of B. John Garrick [4].

… The risks from the nuclear fuel cycle are not included in the safety goals. These fuel cycle risks have been considered in their own right and determined to be quite small. They will continue to receive careful consideration. The possible effects of sabotage or diversion of nuclear material are also not presently included in the safety goals. At present there is *no basis* on which to provide a measure of risk on these matters. *[emphasis added]*

Dealing with the "no basis" perception is one essential point of this report.

1.2 Scenario-Based Assurance in the Safety Domain

In the safety domain, we derive needed assurance from explicit scenario models reflecting system redundancy, diversity, physical margin in component capability, and other system characteristics, and we quantify those properties, up to a point, based on operating history insofar as we have records of it. This includes explicit modeling of initiating event frequencies, explicit modeling of the failures of systems that are supposed to respond to the initiating events, and explicit modeling of the relevant consequences. In principle, the modeling considers the various "treatments" carried out by plant staff to maintain system performance. Moreover, *in the safety domain,* the level of assurance that we require is calibrated to the possible consequence severity *and to the frequency of challenges to the mitigating systems*. If we are sure that the initiating frequency is extremely low (as, for example, "large-break Loss-of-Coolant Accident concurrent with loss of offsite power"), we are less demanding about margin to failure of mitigating systems to meet those challenges. If results of these models provide us adequate assurance that the design is adequate (the likelihoods of the various adverse scenarios are, or can be made, sufficiently low), then attention shifts to *maintaining* the performance we think we need in order to prevent the adverse scenarios.

For a specific top event, such as "core damage," consider the minimal cut set

 $IE^*A^*B^*C^*D^*E^5$ $IE^*A^*B^*C^*D^*E^5$

meaning "Initiating Event IE occurs AND A occurs AND B occurs AND C occurs AND D occurs AND E occurs."

How do we assure that that cut set is unlikely?

IF 5 things have to happen in addition to the initiating event, and if they are independent, and have low individual probability, the cut set will have a fairly low probability. However, there are important cases where assumptions like that are not applicable.

First: a cut set is a conjunction of events. If we could ensure that A^*B can never occur together, then this particular cut set is "false" (in some narratives, impossible), even if both A and B have significant independent probabilities on average. Therefore, the relationship between events in a cut set is important. On the other hand, suppose that A and B are each causable by some identifiable factor X. That is,

$A=A_{ind}+X$, $B=B_{ind}+X$,

where A_{ind} and B_{ind} refer to "independent" failures.

Then, after a bit of Boolean algebra,

⁵ In the present convention, "asterisk" ("*") means "AND" and plus ("+") means "OR."

$A*B=A_{ind}*B_{ind}+X$.

That is, the conjunction A^*B is implied by the single event X. If all of the events in the cut set (A, B, C, A) D, E) can be caused by X, then instead of cut-set occurrence requiring 5 things to happen, it only needs for one thing to happen: X. Some instances of X are shared functional dependences, such as two components relying on the same source of electrical power. Others may be shared harsh environments, or similar vulnerabilities to (say) wearout as a result of design error. Much of the actual work in risk analysis is a search for linkages such as X, which are important because they reduce the redundancy that designers may have been counting on; and once the linkages are identified, proper quantification of the probabilities of those linkages can be important.

The following table appears in "Forging a New Nuclear Safety Construct," by the ASME Presidential Task Force on Response to Japan Nuclear Power Events. [6] The table provides a useful summary comparison of "deterministic" and PRA approaches to safety assessment. The membership of the Presidential Task Force includes some long-time veterans of the nuclear safety establishment, and the table is interesting partly for that reason. In particular, it acknowledges that in the deterministic approach, common-cause failures (CCFs) were "assumed to be precluded by special treatment requirements," while in the PRA approach, they are "probabilistically considered for all equipment based on experience." The deterministic assumption of preclusion of CCF by special treatment is an artifact of early nuclear licensing practice. One reason that CCFs are important in PRA is precisely that many systems in Generation II plants were designed to incorporate identical trains of equipment, consistent with NRC requirements of that era. This feature was deemed to satisfy the single-failure criterion, even though within a PRA perspective, such a configuration might be vulnerable to CCF because of the similarities of the redundant trains.

Table 1 – Comparison of Deterministic and PRA Approaches to Safety

1.3 Characteristics of the Cyber Problem

In the security domain, we need to think about assurance in a different way. In the safety domain, we generally have some objective information about initiating event likelihood, but for reasons to be discussed later, analyzing attack likelihood is problematic. We cannot credibly derive assurance of low scenario likelihood from arguments based on low likelihood of *attack*. We need to derive assurance that adverse *scenario* likelihood is low (attack plus subsequent events) from system characteristics and from the modeled effectiveness of protective measures, and/or the results of security exercises if those results are available. Physical margin in component capability may likewise be inapplicable; we are talking about "systematic" failures, not mechanical or electrical failures.

Moreover, if there are multiple possible scenarios having a particular adverse consequence of concern, we need some assurance regarding prevention of all of those scenarios. We do not prevent the adverse consequence by preventing only some of the possible scenarios.

The above comments can be related to the RIMES approach, to be discussed later. The authors of RIMES argue that assurance is not to be had from estimates of attack likelihood; rather, assurance is to be had from attack *difficulty*.

1.4 Historical Attitudes in Nuclear Safety

The concept of probability does not appear explicitly in most general definitions of margin, but is implied; the point of "margin" is to increase the probability of dealing successfully with contingencies. Put differently: "margin" decreases the probability of crossing a threshold beyond which an adverse outcome will occur. If we are able apply the concept within cyber, the point of increasing "margin" is to increase the probability of not succumbing to some class of possible attacks.

Many would agree with that statement even though some would disagree about whether that probability can be convincingly quantified. Historically, the "margin" concept has been especially popular in contexts within which "probability" is hard to quantify: people who refuse to deal with probabilities are nevertheless willing to discuss "margin." This attitude goes back at least to WASH-740 [7], which states:

As to the probabilities of major reactor accidents, some experts held that numerical estimates of a quantity so vague and uncertain as the likelihood of occurrence of major reactor accidents have no meaning. They declined to express their feeling about this probability in numbers. Others, though admitting similar uncertainty, nevertheless ventured to express their opinions in numerical terms. Estimations so expressed of the probability of reactor accidents having major effects on the public ranged from a chance of one in 100,000 to one in a billion per year for each large reactor. However, **whether numerically expressed or not,** there was no disagreement in the opinion that the probability of major reactor accidents is exceedingly low. **[emphasis added]**

Why did Brookhaven believe that "the probability of major reactor accidents is exceedingly low"? Because:

There are factors both on the side which would lead toward confidence that our "no accident" experience will continue, and on the converse side. On the one hand, we attempt to provide **wide margins of safety** because of our limited knowledge of accident potentials of reactors. The new and glamorous field challenges and attracts the most expert and competent people. The

Government has had and continues to have a substantial safety research program. Experience almost certainly will lead to safer design. On the other hand, since many reactor types are being developed more varied safety problems may exist than would be the case in fewer types. Accident free experience could lead to complacency. Lengthening reactor life could lead to hazards not otherwise encountered (cumulative radiation damage to components). Competitive pressures could furnish incentives to reduce margins of safety. **[emphasis added]**

To repeat: Many people who refuse to deal with probabilities are willing to discuss "margin." We may wish we could deal credibly with probabilities in a security context, but many of us believe that we cannot credibly quantify attack likelihood, and moreover it is not clear that current hazard analyses are as complete as we need them to be. For different reasons, the situation in cyber is analogous to the situation that Brookhaven thought they were facing in safety; hence the present attempt to make sense in terms of margin.

The US Atomic Energy Commission (AEC), which had asked Brookhaven for WASH-740, was not content with Brookhaven's refusal to try to quantify accident likelihoods. Eventually, the US Senate asked for the Reactor Safety Study, WASH-1400 [1], which did quantify accident likelihoods, and appears to have been the first plant-scale study to try to do that in a comprehensive way. However, WASH-1400 did not settle the matter for everyone; even at the turn of the present century, a kind of culture war was ongoing at the US Nuclear Regulatory Commission (NRC) between PRA "believers" and others. (For example, see [8].) In the mid-1990's, NRC Chairman Jackson had formulated a policy of being "risk-informed," [3] but some NRC staff saw "risk-informing" as being a convenient rationale for reducing or eliminating existing safety requirements that had been nominally based on traditional engineering arguments.

It is easy nowadays to dismiss the PRA skeptics. But as we will see later, the Expert Panel on Quantification of Seismic Margins [9] (people who count as "believers") saw value in thinking in terms of a kind of margin that a wide range of practitioners can agree on.

2. Overview of Selected Concepts of Margin

2.1 Common Language Definitions of "Margin"

General definitions of margin and contingency are: [10]

Margin: 4a: something that is over and above what is strictly necessary and that is designed to provide for emergencies: a spare amount or degree allowed or given for contingencies or special situations: …

Contingency: … 2b: a possible future event or condition or an unforeseen occurrence that may necessitate special measures …

Like defense in depth, the "margin" concept conveys the idea of having capability beyond "what is believed to be strictly necessary." In nuclear safety, the point of having margin is to increase the level of assurance that something is safe.

2.2 Nuclear Thermal Hydraulics

Consider the definition of "margin" given in IAEA-TECDOC-1332 [11]:

The safety margin of operating reactors is defined as the difference or ratio in physical units between the limiting value of an assigned parameter the surpassing of which leads to the failure of a system or component, and the actual value of that parameter in the plant. The existence of such margins assure that nuclear power plants (NPPs) operate safely in all modes of operation and at all times. The most important safety margins relate to physical barriers against release of radioactive material, such as fuel matrix and fuel cladding (typical limited values are departure from nucleate boiling ratio — DNBR, fuel temperature, fuel enthalpy, clad temperature, clad strain, clad oxidation), RCS boundary (pressure, stress, material condition), containment (pressure, temperature) and surrounding public dose. In many cases, both the limiting value and actual value are not known precisely, i.e. the safety margin cannot be quantified precisely. Therefore, for practical purposes, the safety margin is usually understood as the difference in physical units between the regulatory acceptance criteria and the results provided by the calculation of the relevant plant parameter.

Figure 1 [from IAEA-TECDOC-1332]. Safety Margins

In the above figure, the y axis (if shown) would correspond to the value of a physical parameter, such as temperature (e.g., temperature of fuel cladding). The "safety limit" would then correspond to a temperature limit whose exceedance would be unsafe. "Safety margin" could be either the difference between a "conservative calculation" and a "Regulatory Acceptance Criterion," or the difference between the Regulatory Acceptance Criterion and a "Value computed by best estimate calculation" plus a correction to allow for uncertainty in that calculation. Either way, the figure aims to suggest that because there is margin, the possibility of exceeding the regulatory acceptance criterion is remote, and the possibility of exceeding the physical safety limits is even more remote, at least in the scenario(s) contemplated in the assessment. The uncertainty is not a precise fixed value; pictorially, the "uncertainty" arrow is drawn to fit on the page, but in practice, it is meant to suggest a characteristic value of a probability distribution.

Operationally, the above concept of margin applies in formal safety analysis, but not necessarily in all real-life situations. The events considered in formal safety analysis are specific and are evaluated according to fixed protocols. Outside of nuclear plant safety analysis, there may be a plethora of potential scenarios to consider, each with its own characteristics and different values of margin. When we conclude that a plant is "safe" based on safety analysis, there is more than one argument being made: the plant has capability for the scenarios considered, including margin, *and that scenario set is an adequate basis for drawing the conclusion.* This is a key difference between PRA and safety analysis. PRA strives for a different kind of completeness, and in practice, this limits the precision (but not necessarily the usefulness) of the kinds of statements that can be based on it.

2.3 Seismic Margins Analysis

In the long run, we will wish to assess whether a given protection scheme is adequate. It seems unlikely that we will be able to convincingly relate any specific protection scheme for a given component to an explicit probability of successful attack on that component; we are more likely to be able to characterize a threshold of sophistication below which an attack will probably fail, given a protection scheme. Actually, we will wish to reason about causing not just an isolated systematic failure, but rather a conjunction of systematic failures corresponding to a particular cut set.

There is a partial precedent for doing something a bit like this: "seismic margins analysis." [9] Put very simply, the idea is as follows. We have a PRA that reflects component failures induced by earthquake, and as usual, we wish to know what risk we are facing. For a mix of reasons, some being mathematical and computational, it is more arduous to quantify a seismic risk model in detail than it is to quantify a socalled "internal events" model in detail (in which component failures are "random," rather than being seismically induced). But useful conclusions can be drawn if, instead of trying to assign probabilities to various seismically-induced failures, we are able to agree on an earthquake level *below which plant-level failure is very unlikely.* This is notionally analogous to characterizing what sorts of attacks a given protection scheme could withstand. The essential difficulty of full quantification does not go away, but we are given a framework within which failure prevention can be discussed meaningfully, even if full quantification of risk is not performed.

In [9], the Panel, in discussing its work, noted that

There is a need to understand how much seismic margin exists. Margin in this context is to be expressed in terms of how much larger must an earthquake be above the safe shutdown earthquake (SSE) before it compromises the safety of the plant.

Furthermore, the Panel went on to recommend ways of thinking about the "margin" concept that they felt would be more helpful than a strictly quantitative approach (explicit quantification of component failure probabilities derived from median seismic capacities):

• Rather than determining a margin earthquake level for a plant, the review methodology determines whether or not a plant has a high confidence of a low probability of failure for an earthquake selected for margin review.

 \bullet

• The Panel believes it is technically more useful and feasible to identify the earthquake motion level at which "there is a high confidence of a low probability of failure" (HCLPF) than to identify the median point of failure. This HCLPF point can serve as a conservative estimate of the

actual earthquake size where the plant (or any component, structure, or system within the plant) has a small probability of being compromised.

• …

Discussing the above summary bullets, the Panel says:

The measure of margin adopted by the Panel is a high-confidence, low probability-of-failure (HCLPF) capacity. This is a conservative representation of capacity and in simple terms corresponds to the earthquake level at which it is extremely unlikely that failure of the component will occur. From the mathematical perspective of a probability distribution on capacity developed in seismic PRA calculations, the HCLPF capacity values are approximately equal to a 95 percent confidence of not exceeding about a five percent probability of failure. …

The Panel expects that a group of engineers would agree on a ground motion level that a component has a high confidence of a low probability of failure [*sic*]; although they are not likely to agree that the confidence can be expressed exactly as 95% -5%. This contrasts with the total lack of agreement which the Panel believes is likely to occur if the median capacity level and associated variabilities were selected as the parameters of interest. This is the reason the panel believes that the HCLPF concept is a more useful way to deal with the whole question of "seismic margins" than an approach using the median fragility values.

Can a seismic HCLPF concept apply within a cyber context? Perhaps…

Table 2 High Confidence Low Probability of Failure (HCLPF)

2.4 The Concept of Margin in the "System-Based Code" (SBC) Concept

Some years ago, analysts in Japan proposed the SBC concept [12] because they recognized that concurrent application of multiple engineering codes and standards within a given problem would lead to unnecessary (and wasteful) safety margin: each engineering code, formulated independently, would factor a presumption of necessary margin into its requirements, and the combination of all these margins would be excessive. The SBC idea was that there could be "margin exchange" among the codes in a given case; one would look at the total margin picture, and trade off different margins to different ends. "Its fundamental philosophy is to pursue optimal global margin distribution by allowing margins exchange among partial codes." An example given in the SBC papers is the recognition that it is possible to compensate for less-desirable materials properties in a given application by inspecting more often. Obviously, inspection does not change materials properties, but inspecting more often means that symptoms of incipient problems are more likely to be detected before the problems occur, which changes the probability of PRA basic events such as "leak" or "rupture." This idea can be applied in the context

of "Reliability Integrity Management" (RIM), [13] an idea in Section XI Division 2 of the ASME Boiler and Pressure Vessel Code, according to which reliability and integrity are "managed" by setting reliability targets and then establishing the monitoring and non-destructive examination schedules and protocols needed to accomplish those reliability targets. From a high level, concepts such as these are far from new, but the idea of establishing such targets at the design stage, and developing observation protocols to match, is still coming into its own.

All "margin" concepts surveyed here are related to tying physical observations to reductions in the probabilities of undesired events. The SBC concept of "margin exchange" proceeds by establishing a target probability (a component unreliability) and then deciding where to get the necessary margin (e.g., excellent materials vs. extra inspection). Based on an example trading material strength against ISI interval, "it was envisioned that margin exchange between material strength and ISI may be possible on a quantitative basis."

2.5 Licensing Modernization Project (LMP)

The Licensing Modernization Project (LMP) [14] is one of the most significant developments in recent years. A comprehensive discussion of it is beyond the scope of the present report; the present subsection focuses on its concept of "margin."

To understand the LMP's concept of margin, we begin with its use of a "frequency-consequence" (F-C) construct. This construct goes back at least to Farmer [15], but has not been used much in nuclear safety since then, although it has seen a lot of use in non-nuclear process industries. The F-C "target" plot below, copied from NEI 18-04 Rev. 1, shows a line in F-C space articulating the "design objective" of ensuring that all scenarios lie below and/or to the left of the line labeled "LBE^{[6](#page-20-1)} F-C Target. Every scenario corresponds to a location on that plot, and each scenario's risk significance is defined by where it lies. The two fuzzy ovals (one red and one green) have been added to the LMP figure to help the reader focus on the "margin" concept.

⁶ LBE stands for "licensing basis event." In the present discussion, we do not need that concept.

Figure 2 F/C Target in the LMP, with notional scenarios added for purposes of the present discussion. [14]

Table 1 Excerpt from the Glossary of [14]

AOO: Anticipated Operational Occurrence; DBE: Design-Basis Event; BDBE: Beyond-Design-Basis Event; TEDE: Total Effective Dose Equivalent; EAB: Exclusion Area Boundary

The fuzziness in each oval is meant to imply uncertainty in frequency and consequences. The green fuzzy oval corresponds to a scenario that is NOT deemed risk-significant; the red fuzzy oval corresponds to a scenario that IS deemed risk-significant.

Apart from illustrating the idea of "margin as relating to reduced probability of adverse consequences," the point of introducing the LMP here is that there is a potentially interesting parallel between the LMP's F-C Target line and the Pareto surface in RIMES. It may prove to be meaningful to discuss "margin" in a RIMES context analogously to the LMP concept.

3. Risk-Informed Management of Enterprise Security (RIMES) and Decision-Making Under Deep Uncertainty (DMDU)

3.1 RIMES

3.1.1 Overview

RIMES [16] was formulated to address physical security, but much of the associated discussion seems to be more broadly applicable. The present application of Top Event Prevention Analysis (TEPA) is aimed more at cyber, but again, much of the discussion is more broadly applicable. A synthesis of these methods, informed by thinking from "Decision Making under Deep Uncertainty" (DMDU),^{[7](#page-22-4)} could be

⁷ DMDU stands for "Decision Making under Deep Uncertainty." In the present discussion, when we speak of DMDU, we are speaking of [17].

very beneficial. This year's task (Application of Multi-Attribute Risk Analysis: Safety, Security, Generation Risk, …) would use such a synthesis if it existed. Accomplishment of the synthesis itself is beyond the scope of this year's task, but such an undertaking might be worthwhile.

3.1.2 Key Points:

- In recent decades, in the reactor safety community, much of risk management has been devoted to implementing what DMDU calls "predict-then-act." That is, risk assessment is used to identify "significant" risks[,](#page-23-1)⁸ and these significant risks become the focus of risk management investments. Per DMDU, this works well when uncertainties are well understood. Risk-informing reactor safety in this way has been generally successful because the uncertainties that exist in reactor safety (as opposed to security) are understood well enough to allow appropriate risk management decisions to be made. However, uncertainties in *security* – in particular, attack likelihood – are argued by some to call for a different approach.
- The "different approach" topic has been recognized for a long time in the security domain, and different workers have proposed different ways to address the issue (or reasons to deny its existence). A feature common to some approaches is the idea that it is essential to allow (at least implicitly) for a range of possible different futures, rather than supposing that we can be confident about which few of those possible futures should be focused on. Not only do we not know what attackers will do: even if we did know their current intentions, they may change their minds. Moreover, there are already many attacker types having different motives and different capabilities. This "range-of-futures" point is a key feature of DMDU, but the potential link between DMDU and security issues appears not to be widely appreciated. (A prominent DMDU example is management of climate change. Many readers will have no trouble understanding the "range-of-futures" idea in that context.) The goal of DMDU approaches is to "satisfice," considering the range of potential futures, rather than trying to optimize within a predict-then-act framework.
- Before risk analysis became as central as it is now, nuclear safety regulation had something in common with "agree-on-decisions." It was believed that analysts could identify severe but credible events that would bound anything that would ever really happen, and if a plant could cope with these bounding events adequately and reliably, it would be deemed safe enough. Curiously, this actually resembles "agree-on-decisions" to some extent: it claims robustness. The claim of robustness was arguably wrong, but at the time, it was sincere. Risk-informing actually drove increased emphasis on "predict-then-act."
- The bounding-event thought process turned out to be trickier to apply than one might think. "Severity" is a multi-attribute thing; it is not adequately captured by a single variable, such as (for example) the rate at which coolant is lost in a loss-of-coolant accident. Hazard identification processes were (and remain) imperfect, so some potential scenarios were missed in the early days. Without a detailed scenario model, It is difficult to think clearly about how to balance various reliability considerations against cost (even for the non-adversarial digital risk problem).
- RIMES pays significant attention to alternative futures. Being focused on security, RIMES' details are more domain-specific than the discussions in DMDU. We are not privy to full details of the analyses on which RIMES publications are based, but it seems that RIMES looks very

⁸ E.g., scenarios that contribute appreciably to the risk metrics.

broadly at physical attack scenarios for security-specific reasons that are analogous to the general concerns that drive the DMDU approach.

- o In particular, RIMES proposes to modify the reactor *safety* risk triplet {scenarios, scenario *frequencies*, scenario consequences} to become {scenarios, scenario *difficulties*, scenario consequences} for purposes of security risk management.
- o RIMES then goes on to argue for a way of setting risk management priorities that resembles to some extent the regret-based approaches that are discussed in DMDU.
- Top Event Prevention Analysis (TEPA) [18] is (like RIMES and DMDU in general) scenariooriented, but proceeds in a different way.
	- *In the safety domain*, TEPA begins with logic-based models of plant risk (PRAs), focusing on specific consequence types (such as "core damage"); such models strive to identify a usefully complete set of scenarios (cut sets) leading to the chosen consequence types, and then looks for risk management approaches that address every "scenario" emerging from the risk model. This sounds like what DMDU and RIMES try to do, but (a) TEPA tries to address "every" scenario (cut set) in the PRA explicitly (discarding only scenarios that can be convincingly shown not to contribute much risk whether or not we allocate resources to their prevention), and *(b) TEPA capitalizes on the point that many scenarios have some elements in common, so focusing on prevention of those elements can be a huge improvement in the efficiency of the overall risk management approach*. For each scenario (cut set), TEPA identifies distinct options for preventing cut set elements that (if implemented) would reduce the implied risk contribution of that cut set to a degree specified in user input; these options are aggregated over all cut sets, leading by construction to identification of a range of risk mitigation approaches that satisfy the user-specified prevention criterion *at the cut set level*. Because TEPA starts at the cut set level, the global (top-event-level) performance of its candidate solutions needs to be tested using the original model; but in practice, for a reasonable prevention criterion, the solutions generally work well, meaning that the risk metrics are acceptable even if significant numbers of components are not credited in the model.
	- *In the security domain*, the mechanics of TEPA can be much the same as they are in the safety domain, but we still need to think harder about the prevention criteria (working with difficulty instead of event probability), and about the actual efficacy of proposed risk management solutions at the plant level. In past applications of TEPA to security (original version of EPRI's HAZCADS [19]), these points were addressed qualitatively based on extensive sensitivity studies (making illustrative assumptions about the efficacy of component protection measures, and understanding the implications of those assumptions at the plant level). In this regard, TEPA resembles the DMDU process of testing candidate strategies against a range of possible futures. But in the security domain, it is not yet clear how to test prevention analysis results by quantifying the sorts of risk metrics that we quantify in the safety domain. In security, we are so far reduced to making qualitative arguments about how difficult we can make successful attacks.

The following subsections furnish additional detail regarding the above points.

"Risk-Informed" Management of Reactor *Safety* **Risk**

The DMDU literature [17] portrays typical PRA applications to safety as being within the "predict-thenact" school of decision-making. By "predict-then-act," they mean

… quantitative predictions of risk (often defined as the predicted probability multiplied by the predicted consequence of an event) to [act by] systematically inform[ing] decisions about the allocation of effort to reduce risk. While often useful when uncertainties are well understood, the approach faces the perils of prediction when uncertainties are deep.

The above is quite a simplification, but it is not wrong regarding many PRA applications in the context of safety. Since the 1981 Kaplan-Garrick paper [20], PRAs have typically referred to risk as a set of "triplets:" {scenarios, frequencies, consequences}. PRA models the scenarios and the consequences, and quantifies the frequencies, allowing throughout for uncertainty. For existing plants, PRA results typically reflect empirically derived frequencies of initiating events, and empirically derived frequencies of specific component failures. In application of PRA results, one rationalizes the allocation of safety resources according to where the risks seem to be significant (e.g., prevention of scenarios that contribute significantly to a risk metric). 10 CFR 50.69 [21] is a specific case of this. The DMDU critique of "predict-then-act" is essentially (in my words) "if you are wrong about what the significant risks are, then you are relatively likely to *regret* your allocation of risk reduction resources." In analyzing the safety of existing light-water reactors, the relevant community of practice considers that for many purposes, the "uncertainties are well understood," so the situation is OK. *However, in security risk, the uncertainties surrounding attack characteristics (likelihood, what gets attacked, etc.), the uncertainties are arguably deep.*

In seemingly distinct but arguably related ways, RIMES, DMDU, and TEPA are all aimed at avoiding "regret."

One reason that the Kaplan-Garrick definition is still widely used is that it is essentially scenario-based. Understanding the scenarios is the beginning of wisdom regarding what might need to be managed, or changed, even if "likelihood" is poorly understood. This would go without saying, except that many people still think of "risk" as "expected consequences," which (if quantified correctly) tells the risk manager how worried to be, but provides no hints as to what to do moving forward. All the methods discussed here respect the Kaplan-Garrick emphasis on scenarios: all of them involve a lot of scenario modeling.

"Risk Informed Management of Enterprise Security" (RIMES)

Wyss and collaborators advocate "Risk Informed Management of Enterprise Security" or RIMES [16]. We do not see "deep uncertainty" explicitly mentioned in their many publications, but they argue persuasively that trying to quantify likelihood will not do a good job of supporting risk management decisions. Along with others cited in the article's references, 9 they discuss the fundamental disagreements regarding quantification of security risk. From the present point of view, the very existence of those disagreements qualifies the subject uncertainties as "deep." From the DMDU book [17]:

… think of rare events like a natural disaster, a financial crisis, or a terrorist attack. These topics are all characterized by what can be called "deep uncertainty." In these situations, the experts do not know or the parties to a decision cannot agree upon (i) the external context of the system, (ii) how the system works and its boundaries, and/or (iii) the outcomes of interest from the system and/or their relative importance (Lempert et al. 2003). Deep uncertainty also arises from actions taken over time in response to unpredictable evolving situations (Haasnoot *et al.* 2013).

Not all aspects of "deep uncertainty" are necessarily present in security risk management, but arguably the "likelihood of attack" aspect tends to qualify as a "deep uncertainty" (if only because there remains

⁹ Some of these references are collected in Appendix A of this report.

significant controversy about it). Operationally, in deep-uncertainty situations, "predict-then-act" decision-making may not work out well; DMDU / robust decision-making is argued to have advantages.

The RIMES approach to addressing this is to replace the classical risk triplet {scenarios, *scenario frequencies*, consequences} with {scenarios, *scenario difficulties*, consequences}. Focusing on scenario difficulty is argued to capture much of why some want to reason with attack likelihood, but without the explicit difficulties that attach to trying to quantify a likelihood. Having taken that step, the RIMES collaboration then plots scenario difficulties and scenario consequences on a figure like the one below, in which each point is a specific scenario, and its coordinates are (difficulty, consequence):

Figure 3 (after RIMES). Each point is a scenario.

The next step in RIMES is to argue for managing risk based on both scenario difficulty and consequence. With not too much difficulty, one can discern in the above figure a rough Pareto surface drawn from the attacker's point of view. Most scenarios in the above figure are suboptimal from an attacker's point of view (thinking only of difficulty and consequences), because for most points, there are scenarios that are easier at the same consequence level, and/or scenarios that are more consequential at the same difficulty. Scenarios that are thus bounded are "inferior" choices for the attacker. Scenarios lying on the surface of noninferior points – the Pareto surface – are not inferior in that sense. At a given difficulty level, the attacker can easily see which scenarios offer the most bang for the buck.

Figure 4 The Essential RIMES Argument.

This is argued in the slide above, which appeared in [24], among other places.

Regret

Because of deep uncertainty (or, for present purposes, the likelihood issue), we cannot credibly calculate the *expected* utility of different decision options in order to pick the best risk management approach (the best pattern of investments in protection^{[10](#page-27-0)}. To address this difficulty, the concept of "regret" is used by some analysts in some situations. The term "regret" has various meanings in the decision-analysis literature; we are talking about the "regret" that arises when we choose a strategy that would be optimal for a specific future, but some other future happens instead, and our strategy performs poorly in that future, so we "regret" choosing the strategy that we chose. For some purposes, what we mean here by "future" is "the attacker's choice of attack." If we think Attack A is likely and allocate all our resources to it, and the attacker chooses Attack B, to which we have paid no attention, we experience regret.

The idea(s) of regret that we may be able to use here are exemplified in the following excerpt from the DMDU book.

RDM [Robust Decision-Making] and other deliberative "agree-on-decision" DMDU methods generally draw robustness criteria from the normative, "agree-on-assumptions" decision-analytic literature. This literature identifies four traditional criteria —called Wald, Hurwicz, Savage, and Laplace—for ranking choices *without*

well-defined probability distributions over future states of the world [emphasis added] (Luce and Raiffa 1957; Schneller and Sphicas 1983). These criteria envision a set of future states of the world f_i and a set of strategies s_i , each with a known

utility, u_{ij} , in each state of the world. If the probability of each future, p_i , were known, the best strategy would be that which yielded the maximum expected

utility, Max_i Σ_i p_iu_{ii}. Lacking such probabilities, Wald selects the strategy that gives the best worse case, $Max_i Min_i(u_{ii})$, and Savage's mini-max regret selects the strategy with the least regret, that is, which deviates least from the best one could choose with perfect information, Min_i Max_i [Max_i (u_{ii}) – u_{ii}]. Both Wald and Savage are conservative in that they attempt to avoid worst cases. In contrast, Hurwicz interpolates between selecting the strategy with the best case and the best worst case, Max_i [αMax_i (u_{ii}) +(1–α)Min_i (u_{ii})], where α represents a level of confidence. Laplace's criterion of insufficient reason assumes equal weighting over all the futures and then selects the strategy that maximizes expected utility. Starr (1962) subsequently proposed the domain criterion, which selects the strategy that has highest utility in the most futures. …

The DMDU book [17] continues its discussion, but for present (security) purposes, the point is simply that "regret" is a topic to be considered. The details of dealing with "regret" will differ in security, where we need to prevent all scenarios, while the reasoning appears to be different for the main example in [17], namely, climate change.

The discussions of RIMES with which we are familiar do not go into enough detail to fully engage the topic of "regret." The RIMES papers make it clear that if you address a scenario that is inferior in difficulty-consequence space, and the attacker picks a noninferior attack at the same difficulty to carry out, you will experience regret. But it is arguably insufficient to say "eliminate (make more difficult) attacks at the Pareto surface." Which attacks should you address first? The ones at the high-consequence end of the Pareto surface?

¹⁰ Calculating *expected* utility would require us to assign probabilities to possible futures.

RIMES-Informed Top Event Prevention Analysis (TEPA)

In management of safety risk, Prevention Analysis takes a somewhat different view, answering a somewhat different question. Prevention Analysis results are not predicated on what we think the scenario frequencies *are*; rather, in the safety context, Prevention Analysis results answer the question "what must be done to reduce accident likelihoods to a level that we *can accept*?" Given scenarios and consequences, how should we allocate safety resources to prevent the top event? For which elements do we need to assure reliability, and what level of reliability do we need to assure, in order to drive down the scenario frequency? Or, in security, what do we do to make successful attack sufficiently difficult? For now, we will stipulate that the worth of not having a specific scenario is tied to the severity of its adverse consequences, such as "core damage" or "large early release" or "large release" or "loss of generation," and ask how we go about assuring adequate prevention of that class of scenarios.

Suppose that a given scenario (cut set) comprises three events A, B, and C, and for simplicity, suppose further that given the consequences of that scenario, we deem "adequate assurance of prevention" to be "satisfaction of the single-failure criterion: prevent at least two of {A, B, C}." Moreover, prevent ALL scenarios (cut sets) to some analogous standard, tied to consequence severity. This sort of reasoning was a pillar of nuclear safety thought before PRA came along. In the above sentences, we elided consideration of the initiating event characteristics, as in fact nuclear safety thought did; scenarios ensuing from all initiating events within the licensing basis needed to be prevented to the standard of the single-failure criterion.

The DMDU literature says a lot about the mechanics of exploring a range of possible scenarios, predicated on ranges of assumptions. Their domains of application are more open-ended than ours: they invest heavily in a thorough investigation of possible futures, including a lot of exploration of the implications of alternative assumptions. But though our problem may seem to be more bounded, a thorough analysis of reactor safety is nevertheless a significant undertaking: we try to develop a model of reactor safety functions that can be used to identify "all"^{[11](#page-28-0)} scenarios leading to specific damage states. In a sense, each cut set represents a family of possible futures, and in reactor safety, we acknowledge a responsibility to prevent all of them if we can. That is the point of reactor risk models: we need to know what range of possible scenarios we need to address in developing mitigation alternatives, and we need to understand how well various mitigation alternatives perform against the whole scenario set. In a way, reactor safety practice is like DMDU in different words; reactor safety takes this path because historically it was considered imperative to be very, very sure that large radiological releases will not occur. Until relatively recently, reactor safety was the only domain that used modeling in this way.

When we apply TEPA to security risk, instead of allocating "reliability" or "quality assurance" or "testing" to various components, we are deciding how difficult we want each cut set to be for the attacker to cause. We are thinking in terms of the RIMES version of the Kaplan-Garrick triplet: {Scenarios, scenario **difficulties**, scenario consequences}. We want to make cut set *elements* difficult enough to cause cut sets to be *very* difficult for attackers to cause, sufficiently that we can live with the resulting (perceived) residual risk.

¹¹ "All" is in quotes because completeness of the scenario set is always an issue at some level. Arguably we can achieve conceptual completeness in the set of success paths, if we consider that specification of a success path implies functional performance of all components that are really in that path, whether or not they appear explicitly; but identifying "all" cut sets is more of a problem. (To identify all cut sets, we need to identify underlying causes of events in order to build a complete logic model, and that's difficult.) We do not need to solve that problem here. The present point is simply that in order to do prevention analysis, we need an adequate understanding of what events combine to yield the adverse consequences, so that we may reason about which of them we need to prevent.

Steps:

- 1. Model the scenarios. (Left-hand side of the figure below.)
- 2. Assign consequences to scenarios.

At this point, we have two elements of the triplet: scenarios and scenario consequences. We have not yet made our allocation decision, so on a RIMES plot, all the scenarios are at (difficulty=0, consequences=scenario-specific consequences). This is what's shown on the left side of the figure below.

- 3. Bin the scenarios by consequence.
- 4. Decide what level of "risk" you can live with for each consequence type.
- 5. Allocate reliability / difficulty targets to elements of the model according to what it takes to drive high-consequence scenarios down to a frequency / difficulty that you can live with. (Right-hand side of Figure 5)

Figure 5 Steps in RIMES-Informed TEPA.

On the right-hand side of the figure, each scenario has been allocated prevention resources to achieve some desirable level of difficulty given the consequence type that the scenario causes. Most reactor safety people would consider "Large Early Release" to be the most severe of the four consequence bins shown, because "Early" means "before effective evacuation of the public." Accordingly, based on the prevention criterion, the easiest of those scenarios is harder than the easiest of any of the other consequence bins. The "Large Release" bin is next; since Fukushima, there has been increased attention worldwide to releases that are not "Early," because at Fukushima, such releases caused considerable distress by forcing evacuation.^{[12](#page-29-0)} The "Core Damage" bin is next: scenarios in this bin essentially destroy the plant, but cause neither fatalities nor evacuation. Finally, Loss of Generation is shown notionally.

 12 Thousands of people are still displaced as a result of the releases at Fukushima. Though not much discussed, some investigators believe that much of the material released actually blew offshore, and the consequences could have been much worse than they were.

What about further improvements?

Beginning with the right-hand portion of Figure 5, RIMES would identify the Pareto surface and look for improvements in the scenarios lying at that surface. We have context-specific things to consider that are not mentioned in the RIMES papers: for example, large early releases need to be very, very unlikely. The discussion above tacitly assumes that in the process of TEPA itself, we imposed category-consequencebased prevention criteria on each scenario category.

We have not yet applied the following observation, but it is interesting to note that a RIMES plot (such as Figure 3) can be put into correspondence with a LMP F-C plot if one rotates the RIMES plot 90° to the right, and then draws the Pareto surface on the RIMES plot. The F-C line on the LMP figure represents a facility-level objective; the Pareto surface on the RIMES plot represents a current situation. If we knew how much difficulty is enough, we could draw a surface on the RIMES plot corresponding to a facilitylevel objective, and reason about the implications of that analogously to LMP.

Section 4 of this report does a case of part of Figure 5 (extra prevention applied to LERF) with a view to demonstrating a computationally efficient way of actually doing the analysis.

4. Conditioning a Prevention Analysis of Large Early Release on an Existing Prevention Analysis of Core Damage

This section presents the results of a computational experiment to illustrate conditioning of a LER Prevention Analysis on a previous [22] Prevention Analysis of Core Damage.

4.1 Background

When Top Event Prevention was first formulated in the 1980 's, 13 13 13 the point was to identify and rationalize the set of structures, systems, and components that needs special treatment, in order to make a licensing safety case. The top event analyzed was "radiological release from a waste-handling facility." The idea then was to simply analyze the results of a model for radiological release, impose a prevention criterion based on regulatory considerations, obtain a collection of Prevention Sets, and choose one for implementation based on considerations of efficiency.

For many problems, determining a complete set of Prevention Sets is a significant problem computationally. As explained in previous reports, the calculation is equivalent to computing the logical AND of a complete collection of cut-set-level prevention expressions, one expression per minimal cut set. Many facility analyses generate very large numbers of cut sets, so this computation is very difficult. Pending some disruptive discovery in methods for efficient Boolean processing, the calculation is always done stepwise. To see why, suppose we have 100000 cut sets, and a prevention expression for each cut set involving 3 terms. We need to compute the result equivalent to formulating the logical AND of all of these cut-set-prevention expressions, expanding that quantity, and performing Boolean reduction of it to yield the Prevention Set expression. But trying to do that by brute force yields an intermediate stage in the calculation having a number of terms on the order of 3^{100000} , which is impractical. If, instead, we compute the logical AND of the first two prevention expressions, reduce the result, compute the AND of that result with the next prevention expression, and so on, we can at least make a good deal of progress.

¹³ The method was formulated in 1988, but the first paper to use the name "Top Event Prevention Analysis" [18] did not appear until the 1990's.

In short: in problems of practical significance, we always proceed in stages. But even if we do that, the problem is a very large one.

Nowadays, there is great interest in understanding not only a top-level risk metric such as "radiological release," but also subsidiary metrics. Formally, one could argue that the only public safety aspect of nuclear plant safety is "release," and modeling and regulation should focus on that outcome; but such a perspective reflects excessive focus on the nominal top result of a very complicated model that may be difficult to trust. In order to trust a large-scale model, we need to understand something about intermediate-level results, not just what the model says about "release." For example, in the case of Generation II light-water reactors, it is nowadays customary in the US to consider both the frequency of core damage, and the frequency of large "early" release. Going beyond that: in order to think usefully about defense in depth, one needs to consider not over-relying on any one function (much less any one system, much less any one component). This has always been a feature of nuclear safety thought, and in recent years, it has received increasing attention in the formulation of safety cases for advanced designs. For example, it is prominently discussed in NEI 18-04 [14].

Let us return now to the matter of 3^{100000} , the number of terms in the intermediate calculational stage of a brute-force Prevention Analysis exercise on a sizable model result. All of these solutions nominally satisfy the cut-set-level prevention criteria; but it is not clear whether they will all satisfy the other *desiderata* implied in the previous paragraph. If we require balance in the performance metrics of various systems and components (essentially, if we require defense in depth), the solution space will become smaller: easier to handle and easier to understand.

Can we accomplish this reduction *a priori*, as opposed to generating all of the top-level Prevention Sets first, and then somehow pruning the set?

Yes. The next subsection illustrates the calculational process on the sizable problem we have been working with (the Grizzly Gulch Generating Station PRA). This example does not go into the details of defense in depth, but shows how to *condition* the Large Early Release result on the Core Damage result. Addressing defense in depth from this point of view would be a handful of much smaller problems, which would be interesting from a design / systems-engineering perspective but less challenging computationally.

4.2 Computation

One simple approach to getting a Prevention Set for large early release is simply to obtain the cut sets for large early release, and run the prevention analysis process on that collection of cut sets. That approach is conceptually simple but not necessarily best, and in some cases, it is not even the most computationally efficient approach. Consider instead:

- obtaining Prevention Sets for core damage,
- choosing one of those Prevention Sets based on efficiency, cost, and so on, and then
- conditioning the LERF prevention analysis on that Prevention Set.

This section briefly presents such a calculation.

Formally, it is not wrong to do the core-damage and large-release problems as if they were independent, but neither does it automatically yield an efficient joint solution: at the end, we probably need to choose a CD Prevention Set and an LER Prevention Set for implementation, and it is desirable for the LER

Prevention Set to be properly conditioned on the CD Prevention Set. That is: the LER Prevention Set does not need to replicate the capabilities that are already built into the core-damage Prevention Set, but it needs to fill whatever gaps the core-damage Prevention Set has. Doing the problems independently, we may not achieve that outcome, and even if we did achieve it, that fact might be tedious to demonstrate. Implementation of independently derived Prevention Sets will accomplish the prevention objectives, but not necessarily in an efficient way.

Steps Actually Done

(Variations are possible, but this is what was done in the present illustration)

Given:

- Cut sets for Core Damage.
- Cut sets for Large Early Release (core damage cut sets that have been binned into a LER release category)

Do the following:

1. Obtain a Prevention Set for core damage in the usual way. Choose a Prevention Set on which to condition the LER analysis.

In the present example, Level 2 Prevention Analysis was done for core damage, and the selected Prevention Set was further augmented through the use of importance measures. This augmented Prevention Set was the starting point for the LER analysis.

- 2. Prepare input for a targeted version of the "Preventalation" process.
- In general, Prevention Analysis works from a user-formulated prevention criterion of some kind, writing a prevention expression for every cut set (we refer to this as "Preventalation"), and computing the logical "AND" of the resulting cut-set-level prevention expressions.^{[14](#page-32-0)} This yields a prevention expression covering the whole model result: we prevent the top event by preventing every minimal cut set to some desired level. For example, in the present case, the core damage Prevention Set was obtained by finding Level 2 Prevention Sets (sets of basic events whose protection prevents at least two events in every cut set), and choosing one of those Prevention Sets. As explained in the previous milestone report, the chosen Prevention Set was then enhanced by selective application of an importance measure. This enhanced Prevention Set was used in the previous milestone report to obtain Generation Risk Prevention Sets by conditioning the Generation Risk Prevention Analysis on a chosen level 2 Prevention Set for Core Damage. A generally similar thing was done here to find LER Prevention Sets.
- It is often desirable in Prevention Analysis to steer the calculation. One way to steer it is to tell the calculation what events it should NEVER take credit for preventing, and what basic events it should ALWAYS take credit for preventing. This capability was motivated by other considerations, but is highly useful in the present example. In the Preventalator used here, the "ALWAYS" option leads to formulation of cut-set-level prevention expressions formulated in terms of a prefactor containing the logical AND of the "ALWAYS" events appearing in that cut set, ANDED with a factor containing the rest of the logic needed to prevent the cut set to the desired level. For example, suppose we are given cut set $A^*B^*C^*D$, and wish to prevent it to Level 3. The corresponding prevention expression is

 $A*B*C+A*B*D+A*C*D+B*C*D.$

But if we wish ALWAYS to take credit for A*B, then the resulting cut set level prevention logic looks like this:

Prefactor in the prevention expression=A*B.

Rest of the prevention expression=C+D.

Level 3 Cut Set Prevention= Prefactor * The Rest = $(A*B)*(C+D) = A*B*C+A*B*D$.

¹⁴ Here, this process was performed using Top Event Prevention Analysis software. [23]

This approach to Preventalation was done originally for scrutability and efficiency, but for present purposes, it is useful for another reason. In order to condition our LER analysis on a Core Damage result, we "ALWAYS" all of the basic events in the selected core damage Prevention Set; then, when we develop Level 3 prevention expressions from the LERF cut sets, the factors representing events chosen for CD prevention appear in prefactors, while events that are NOT part of the CD Prevention Set appear in "The Rest" factors. We then do the *conditional* Prevention Analysis without the prefactors, which reduces the problem somewhat. We skip over the prefactors, and focus on "the rest" for each cut set. We have several possible cases for each cut set:

- The Prefactor alone satisfies the prevention criterion, and "The Rest" is vacuous for the subject cut set;
- The Prefactor satisfies part of the prevention criterion, and "The Rest" is correspondingly reduced from what would be needed without the prefactor;
- The Prefactor contains none of the cut set variables, and is vacuous, and "The Rest" is the full prevention expression for the cut set.

An exception occurs when the Level 2 Prevention Criterion is satisfied for the Core Damage analysis, but the Level 3 Prevention Criterion cannot be satisfied for the LER analysis: there is a two-element cut set, but we need to prevent three elements. The Preventalator behaves differently in this case. This turns out to be a feature, not a bug: it flags the cut sets for which Level 2 could be achieved, but Level 3 could not, which is a very interesting set of events to flag.

3. The LER Prevention Sets conditioned on the CD Prevention Set are obtained by skipping the prefactors, and handling "the rest" factors as if they were full prevention expressions. This was done using commercial Top Event Prevention software [23]. The resulting conditional Prevention Sets, ANDed with the CD Prevention Set chosen in Step 1, are LERF Prevention Sets obtained through a process that is less intensive computationally than a full LER prevention exercise starting from scratch.

In fact, this problem is too large to permit finding "all" of the Prevention Sets, even just for the LER problem. We were working with nearly 150000 cut sets for the LER problem (still just a truncated set), and for those cut sets, there are potentially hundreds of millions of Prevention Sets. In the present case, the calculation was allowed to process one cut set at a time until the problem became too large. The resulting Prevention Set expression up to that point prevents the cut sets processed up to that point. That intermediate Prevention Set result was truncated to keep only the smaller (presumptively more desirable) Prevention Sets, and the calculation proceeded from that point to factor in the rest of the cut set prevention expressions. This manual truncation process needed to be applied twice in the present example.

The table in Appendix B shows, for each basic event variable, whether it appeared in the original conditioning Core Damage Prevention Set and whether it appears in the LER Prevention Set chosen from the sets emerging from the above steps.

The "exception" mentioned above is also shown in this table; some of the variables appear in both the CD Prevention Set and the conditional LER Prevention Set. How does this occur, if the LER Prevention Set is properly conditioned on the CD Prevention Set? Consider a two-element cut set. If we run Level 2 Prevention Analysis on the CD expression, both elements in that two-element cut set will be required in every CD Prevention Set. But when we try to run Level 3 Prevention for LER, we will find that the Level 3 criterion is not satisfiable. The present version of the Preventalator, not foreseeing the present application, does not apply its usual "prefactor" processing, and simply writes the whole prevention

expression including the AND of the two cut set events that are there, ANDed with a "slack variable" standing for the third basic event whose prevention would be needed to truly satisfy Level 3 for that cut set.

Useful information is encoded in the slack-variable names. After "SLK," standing for "Slack," the number in the middle of the name is the number of the cut set that was "short" of the number of variables needed to satisfy the prevention criterion; and "SHN" means that that cut set was **SH**ort by N events. If we wish to implement the prevention criterion literally, we need to go to that cut set, and figure out what, if anything, to do. For example, for a specific cut set, we might find that the two events that are there are highly preventable, so having only two of them is not really a problem.

The first slack variable listed in Table 3 is associated with cut set #1105,

$$
Q - 589 \times Q - 608 \times Q - 48 \times Q - 137.
$$

Consulting the listing in Appendix B, we find that Event 589 is the Steam Generator Tube Rupture Initiating Event, and Event 608 is a flag used by the PRA software to disposition this cut set into the "LER" category. Neither is being counted in prevention. That leaves two

Table 2 Slack Variables Appearing in LER Prevention Sets

Useful information is encoded in the slack-variable names. After "SLK," standing for "Slack," the number in the middle of the name is the number of the cut set that was "short" of the number of variables needed to satisfy the prevention criterion; and "SHN" means that that cut set was **SH**ort by N events. If we wish to implement the prevention criterion literally, we need to go to that cut set, and figure out what, if anything, to do. For example, for a specific cut set, we might find that the two events that are there are highly preventable, so having only two of them is not really a problem.

5. Summary and Next Steps

In the previous Milestone Report [22], we illustrated the potential value of considering multiple attributes (core damage risk and generation risk) in a Top Event Prevention Analysis. The attributes addressed in that work were high-level metrics addressing different *values* (safety and production); the point of

addressing them together was to show how to benefit from synergy by choosing to protect equipment promoting both objectives. In the present report, we have done a computationally similar thing but applied it to distinct metrics that both applied to "safety" (instead of safety and production): we developed a prevention analysis of "large early release" conditional on a Prevention Set selected to prevent core damage.

Since public safety is not directly threatened by core damage without release, why not simply focus on large early release to begin with? The modern view of risk management (e.g, "risk-informing" [3]) discourages relying on a single top-level metric, but encourages applying diverse perspectives including defense in depth. A hierarchical application of TEPA could lead to good performance at multiple levels of the performance hierarchy: component, system, function, and plant levels. The technique demonstrated in Section 4 is one step in demonstrating fulfillment of the various safety objectives considered in the Licensing Modernization Project [14], which is likely to play an important role in licensing of advanced designs. We have not yet tested this on a full suite of lower-level (e.g., functionlevel) Prevention Sets, but plan to explore it in the near term. In general, adding constraints to a Prevention Analysis problem reduces the size of the solution space, which can be a good thing, but it is also possible for added constraints to preclude desirable solutions. Learning how to do this well has significant potential value to applications of the LMP.

While the Section 4 demonstration addressed computational aspects of the analysis, it did not solve the problem of setting priorities in actual prevention. We still do not have a better concept of margin against cyberattack than the concept of "Prevention Level" (number of layers of components protected). However, we are planning tabletops in FY 25 intended to do a better job on that issue. To support those tabletops, we have

- collected useful background information about the concept of margin, which should be valuable in planning the tabletop exercises;
- identified a potentially useful path forward for cyber risk management through application of "decision-making under deep uncertainty" (DMDU). This family of methodologies has certain high-level compatibilities with TEPA and with RIMES.

In planning the tabletops, while recognizing the special issues raised by "deep uncertainty," we arguably need to fall back on *processes* that responsibly develop consensus regarding the effectiveness of possible approaches to protection of target sets. Processes associated with some of the historical "margin" concepts may add value. Certain topics in seismic risk analysis could be useful in reasoning about attacks: not because earthquakes are closely analogous to adversarial attacks, but because of the thought processes applied by the Seismic Margins [9] team and the Senior Seismic Hazard Analysis Committee [24] in setting risk management priorities in very uncertain situations *without detailed quantification of event probabilities*. These include:

- the point that a community of experts found it useful to agree on an earthquake magnitude at which they had a "High Confidence of Low Probability of Failure" (HCLPF), rather than trying to quantify explicit probabilities of adverse outcomes;
- their way of thinking about scenarios as being characterized (for some purposes) in terms of the difficulty of accomplishing the single most difficult element of the scenario. (In the seismic context, this meant focusing on the scenario element having the highest seismic capacity, requiring the largest earthquake to cause its failure. This goes beyond "number of elements needing to be compromised" to focus on "difficulty of compromising the most resistant element.")
- in speaking about Probabilistic Seismic Hazard Analysis (PSHA), the Senior Seismic Hazard Analysis Committee (SSHAC) had an interesting take on the form of the desired result, and the reason why the result needed to have that form:

The most important and fundamental fact that must be understood about a PSHA is that the objective of estimating annual frequencies of exceedance of earthquake-caused ground motions can be attained only with significant uncertainty. Despite much recent research, major gaps exist in our understanding of the mechanisms that cause earthquakes and of the processes that govern bow an earthquake's energy propagates from its origin beneath the earth's surface to various points near and far on the surface. The limited information that does exist can be-and often is-legitimately interpreted quite differently by different experts, and these differences of interpretation translate into important uncertainties in the numerical results from a PSHA.

With the above in mind,

SSHAC believes that the following should be sought in a properly executed PSHA project for a given difficult technical issue: (1) *a representation of the legitimate range of technically supportable interpretations among the entire informed technical community,* and (2) the relative importance or credibility that should be given to the differing hypotheses across that range. As SSHAC has framed the methodology, this information is what the PSHA practitioner is charged to seek out… *[emphasis added]*

The emphasis on the *range* of interpretations and the cautious approach to assigning "credibility" to various futures, while not identical to the approaches in RIMES and DMDU, is analogous in spirit to "agree on decisions." If we can achieve this in the tabletops, they will be a success.

6. References

- 1.Reactor Safety Study, WASH-1400 (NUREG-75-014), US Nuclear Regulatory Commission, 1975. This study was conducted by a team led by Prof. Norman Rasmussen of MIT, and is sometimes called the "Rasmussen Report."
- 2. 60 FR 42622, Use of Probabilistic Risk Assessment Methods in Nuclear Regulatory Activities, Final Policy Statement.
- 3.Jackson RI speech [Speech by NRC Chairman Jackson to the Plant Life Management and Plant Life Extension International Conference and Exhibition, December 8, 1997 (No. S-97-26).]
- 4.B. John Garrick, "Principles of Unified Systems Safety Analysis," Nuclear Engineering and Design **13,** pp. 245-321 (North-Holland Publishing Company, 1970).
- 5. Safety Goals for the Operations of Nuclear Power Plants; Policy Statement; Republication (51 FR 28044, 8/4/86; 51 FR 30028, Published 821/86).
- 6. ASME Task Force Forging a New Nuclear Safety Construct / The ASME Presidential Task Force on Response to Japan Nuclear Power Plant Events, ASME Presidential Task Force: Nils Diaz, Regis Matzie, Kenneth R. Balkey, John D. Bendo, John C. Devine Jr., Romney B. Duffey, Robert W. Evans, Thomas R. Hafera, James A. Lake, David E. W. Leaver, Robert Lutz, Jr., Roger J. Mattson, Richard R. Schultz, J. Robert Sims, Douglas E. True, John M. Tuohy Jr. (ASME, 2012).
- 7. WASH-740, "Theoretical Possibilities and Consequences of Major Accidents in Large Nuclear Power Plants / A Study of Possible Consequences if Certain Assumed Accidents, Theoretically Possible but Highly Improbable, Were to Occur in Large Nuclear Power Plants," WASH-740 (March 1957), performed largely by staff of Brookhaven National Laboratory.
- 8. "Report on Interviews and Focus Group Discussions on Risk-Informed Activities in the NRC Reactor Program," Evelyn Wight (WPI), Leila Peterson (WPI),Mark Caruso (NRC/NRR), August Spector (NRC/NRR), Stuart Magruder (NRC/NRR, Robert Youngblood (ISL, Inc.), Kim Green (ISL, Inc.), Attachment 2 to Memo from Gary Holahan to NRR Executive Team and NRC Deputy Regional Administrators, "Results of Internal Focus Group Discussions and Interviews Regarding

the Use of Risk-Informed Regulatory Approaches in the Reactor Program," ML022460161 (NRC, August 30, 2002).

- 9. Seismic Margins. a report by the "Expert Panel on Quantification of Seismic Margins" (An Approach to the Quantification of Seismic Margins in Nuclear Power Plants, R. J. Budnitz, P. J. Amico, C. A. Cornell, W. J. Hall, R. P. Kennedy, J. W. Reed, M. Shinozuka, NUREG/CR-4334 (USNRC, 1985)),
- 10. Webster's Third New International Dictionary / Unabridged (1961)).
- 11. IAEA-TECDOC-1332, Safety margins of operating reactors / Analysis of uncertainties and implications for decision making (IAEA, 2003)
- 12. Yasuhide Asada, Masanori Toshino, and Masahiro Ueta, "System-Based Code Principal Concept," Proceedings of ICONE10, 10th International Conference on Nuclear Engineering, April 14-18, 2002, Arlington, Virginia, USA (ICONE10-22730).
- 13. Boiler and Pressure Vessel Code Section XI Division 2, Reliability Integrity Management (American Society of Mechanical Engineers).
- 14. Risk-Informed Performance-Based Technology Inclusive Guidance for Non-Light Water Reactor Licensing Basis Development, NEI 18-04 Rev. 1 (Nuclear Energy Institute), available from NRC's ADAMS (ML19241A472).
- 15. F. R. Farmer, Siting criteria A new approach (SM-89/34), in Proceedings of a Symposium / Vienna, 3-7 April 1967, on Containment and Siting of Nuclear Power Plants (International Atomic Energy Agency, Vienna, 1967).
- 16. For a recent reference on RIMES, see Gregory D. Wyss & Adam D. Williams (2023), "Possible Does Not Mean Useful: The Role of Probability of Attack in Security Risk Management," Nuclear Science and Engineering, 197:sup1, S80-S94, DOI: 10.1080/00295639.2022.2129224. See also "Risk Informed Management of Enterprise Security (RIMES)," or, "Why computing security risk based on a probability of attack is possible, but is likely not useful for Risk Management," Presented at PSAM 16 – June 27-July 1, 2022 – Honolulu, HI (Gregory D. Wyss).
- 17. **Making Good Decisions Without Predictions / Robust Decision Making for Planning Under Deep Uncertainty**, Published Feb 28, 2013 by Robert J. Lempert, Steven W. Popper, David G. Groves, Nidhi Kalra, Jordan R. Fischbach, Steven C. Bankes, Benjamin P. Bryant, Myles T. Collins, Klaus Keller, Andrew Hackbarth, et al. (downloaded from RAND / Research & Commentary / Research Briefs / on March 26, 2024, https://www.rand.org/pubs/research_briefs/RB9701.html).
- 18. R. W. Youngblood and R. B. Worrell, "Top Event Prevention in Complex Systems," Proceedings of the 1995 Joint ASME/JSME Pressure Vessels and Piping Conference, PVP-Vol. 296, SERA-Vol. 3, "Risk and Safety Assessments: Where Is the Balance?" July 1995 (The American Society of Mechanical Engineers, New York, New York 10017, 1995).
- 19. HAZCADS: Hazards and Consequences Analysis for Digital Systems. EPRI, Palo Alto, CA: 2018.3002012755.
- 20. "On The Quantitative Definition of Risk," Stanley Kaplan and B. John Garrick, Risk Analysis, Vol. **1**, No. 1, 1981.
- 21. Code of Federal Regulations, Title 10, Chapter 1, Part 50, §50.69, Risk-informed categorization and treatment of structures, systems and components for nuclear power reactors.
- 22. R. Youngblood, D. P. Blanchard, and M. Diaconeasa, "Complete Baseline Analysis of a Plant Risk Model," INL/RPT-24-77663 (April 2024).
- 23. D. P. Blanchard and R. B. Worrell, Top Event Prevention (TEP) Instructions, Applied Reliability Engineering Inc. and Logic Analysts Inc., 2012.
- 24. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, Main Report, Prepared by Senior Seismic Hazard Analysis Committee (SSHAC), R. J. Budnitz (Chairman), G. Apostolakis, D. M. Boore, L. S. Cluff, K. J. Coppersmith, C. A. Cornell, and P. A. Morris, NUREG/CR-6372 (USNRC, 1997).

Appendix A: Attack Likelihood

This appendix points to problems associated with trying to treat attack likelihood analogously to the likelihoods of PRA initiating events. A particularly useful summary has been provided by Sandoval [1], who provides numerous references [2-8]. More recently, discussions of RIMES [9, and references contained therein] have built on that history.

Key parts of Sandoval's discussion are:

Professor Norm Rasmussen and a team demonstrate key PRA principles while conducting the reactor safety study (WASH-1400).[2] A modified version of the societal risk model in WASH-1400 was proposed for nuclear safeguards. Known as the ERDA-7 proposal,[3] it proposed portraying safeguards risk as:

 $Risk = F x P x C$

where

F is frequency of occurrence of an attack at a nuclear site, P is the probability the attack is successful,

C is the consequence of the attack.

Issues with that proposal included the following.

- 1976 A basic assumption in PRA is that probability of occurrence is based on failures that are random in nature allowing use of appropriate mathematical models – In the case of deliberate human action such an assumption is surely not valid. [4]
- 1982 Errors caused by the assumption that attempt frequencies (probability of occurrence) are random. [5]
- 2008 Intrinsic subjectivity, interdependency, and ambiguity of threat (Likelihood of Occurrence). [6]
- 2010 The (National Academy of Science) committee advises against the use of probabilistic risk assessment (PRA) in designing security for the DOE nuclear weapons complex. [7]
- 2013 Attack frequencies estimates, even in the aggregate, have too much uncertainty to be useful. [8]

More recently, Wyss and collaborators have discussed this issue as part of their development of "Risk-Informed Management of Enterprise Security," which instead of using attack *likelihood* and scenario consequences, uses attack *difficulty* and scenario consequences. See for example [9].

Appendix A References:

- 1. Joseph Sandoval, History of Vulnerability Assessments in the U.S. Nuclear Program, SAND2014-1003C/SAND2020-8266 V/SAND2020-6566 C
- 2. Reactor Safety Study, WASH-1400 (NUREG-75-014), US Nuclear Regulatory Commission, 1975. This study was conducted by a team led by Prof. Norman Rasmussen of MIT, and is sometimes called the "Rasmussen Report."
- 3. W. M. Murphey, T. S. Sherr, C. A. Bennett, Societal Risk Approach to Safeguards Design and Evaluation, ERDA-7, 1975.
- 4. 1 Rasmussen, Norman C., Probabilistic Risk Analysis Its Possible Use in Safeguards Problems, Proceedings of the INMM, 1976.
- 5. J. M. Richardson, Comprehensive Safeguards Evaluation Methods and Societal Risk Analysis, SAND82-0366, Sandia National Laboratories, 1982.
- 6. L. A. Cox, Jr., Some Limitations of "Risk = Threat \times Vulnerability \times Consequence" for Risk Analysis of Terrorist Attacks, Risk Analysis, Vol. 28, No. 6, 2008.
- 7. Committee on Risk-Based Approaches for Securing the DOE Nuclear Weapons Complex, Understanding and Managing Risk in Security Systems for the DOE Nuclear Weapons Complex, 2010.
- 8. Aaron Clauset and Ryan Woodard, Estimating the Historical and Future Probabilities of Large Terrorist Events, The Annuls [sic] of Applied Statistics, 2013.
- 9. RIMES See Gregory D. Wyss & Adam D. Williams (2023) / Possible Does Not Mean Useful: The Role of Probability of Attack in Security Risk Management, Nuclear Science and Engineering, 197:sup1, S80-S94, DOI: 10.1080/00295639.2022.2129224. See also their many references.

Appendix B: Prevention Sets

Core Damage Prevention Set Compared with Large Early Release Prevention Set The table below lists basic events in the GGGS PRA model and indicates whether they are in the Core Damage Prevention Set chosen as the basis for the Large Early Release Prevention Set, or in that Large Early Release Prevention Set. The 106 Variables needed in EVERY Level 3 LER Prevention Set are indicated in the next-to-last column. The shortest LERF Prevention Set obtained needs one variable in addition to the 106 needed in every LERF Prevention Set (#404).

The importance-measure-based method for enhancing Prevention Sets shown in the previous report was done for the selected Core Damage Prevention Set but has not been carried out here for the LER Prevention Set. The point of the present analysis has been to stress the relative ease of starting from a Prevention Set addressing one attribute in developing Prevention Sets addressing other attributes.

