Carnegie Mellon University
Software Engineering Institute

AGGREGATE INDICATOR MEASUREMENT METHOD CHARACTERIZATION

SEI CERT Division

National Insider Threat Center

October 2020

DOI: 10.1184/R1/13138580

1 Introduction

In this report, we characterize the primary methods for measuring the probability of insider attack by aggregating insider threat indicators. We focus specifically on the methods that were (1) widely reported to be successfully used on the insider threat problem and (2) incorporated into prominent insider threat behavioral analytics tools. In addition, we emphasize methods that are consistent with baseline tools and capabilities observed in mature insider threat programs. The SEI's CERT Division is planning a subsequent report about more sophisticated behavioral analytic techniques that can play an important role in future insider threat mitigation. This future report will consider recent work by the SEI to characterize machine learning in network security[1] and by Gartner to define AI for IT Operations (AIOps) [Rich 2019].

This report builds on a framework for understanding *data model mapping transforms* described in previous work. The framework maps data to observations to indicators to behavior; these concepts are succinctly characterized by the original authors of the paper *Identifying At-Risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats* [Greitzer 2009]:

Data – "Directly available information"

Observation – "Inference from data that reflects a specific state"

Indicator – "Action/event as evidence of precursor to inferred behavior"

Behavior – "Sequence of actions associated with a purpose"

[1] B. Cohen, J. Fallon, A. Horneman, "Machine Learning in Network Security," Software Engineering Institute Special Report CMU/SEI-2020-SR-0252, April 2020.

CERT researchers describe specific transform functions between each successive concept as follows:

> *"... they can be thought of as*
>
> $O = f(D) + \epsilon$
>
> $I = f(O) + \epsilon$
>
> $B = f(I) + \epsilon$
>
> *where D is the data source(s), O is the observation(s) transformed from the data, I is the indictor(s) transformed from observation(s), and B is the behavior(s) transformed from the indicator(s). The epsilon (ϵ) associated with each transform represents the amount of uncertainty in the transforms. This uncertainty could represent the number of false positives or false negatives introduced by the transform. Overall, it represents some amount of risk associated with the analytics used by the insider threat program associated with these data mapping transform functions."*[2]

While the previous CERT work focused on the transform from observation (O) to indicator (I) and the associated cost matrix for methods supporting that transformation, this report focuses on the transform from indicator (I) to behavior (B). We are interested in the behaviors that reflect an increased probability of insider attack; so, measuring combinations of indicators involves quantifying the probability of attack associated with those combinations. Of course, the behavior dimension can be modeled explicitly as malicious behavior patterns that involve actions that inflict organizational harm (e.g., through threat scenarios or using behavioral modeling techniques). Such approaches are not central to current operational insider threat behavioral analytics, but their importance for this domain suggests that we consider them in our subsequent report. In this report, we restrict our attention to the mapping of indicators directly to the probability of insider attack.

Section 2 of this report describes the overall structure for classifying methods supporting the I to B transform and the associated cost matrix for that transform. Section 3 elaborates on that structure for more rigorous, quantitative methods, positioning representative methods and applications within that structure. In Section 4, we summarize the primary contributions of the report.

---

[2] CERT National Insider Threat Center. Insider Threat Indicator Cost Matrix. November 2019.

2 Landscape of Aggregate Indicator Measurement

Figure 1 shows the landscape of aggregate indicator measurement methods along two dimensions. The y axis encompasses the range from methods based on classical statistical inference to methods based on Bayesian inference. Classical methods interrelate variables (indicators) using classical statistical methods, such as statistical correlation, while Bayesian interrelationships are characterized through conditional probabilities. In Figure 1, we focus on rigorous, quantitative methods, which are usually based on ratio-scale measurement, rather than predominately qualitative methods, which are often based on an ordinal scale.

The x axis encompasses the range from subjective, human-judgment-established methods to objective, empirical-data-established methods. Both classical and Bayesian approaches can also be characterized by whether they are developed based on human judgment, empirical data, or a combination of both. Once either the classical or Bayesian model is *established*, these models can be *used* to make decisions based on empirical data generated as organizational systems operate. As already mentioned, these methods are built on methods that support the data to observation, and the observation to indicator transforms. Other artificial intelligence (AI) and machine learning (ML) techniques, such as neural networks [Yuan 2018] or hidden Markov chains [Tabish 2016] (the subject of our future report), can be used in more sophisticated approaches and as input to empirical Bayesian or Regression approaches.

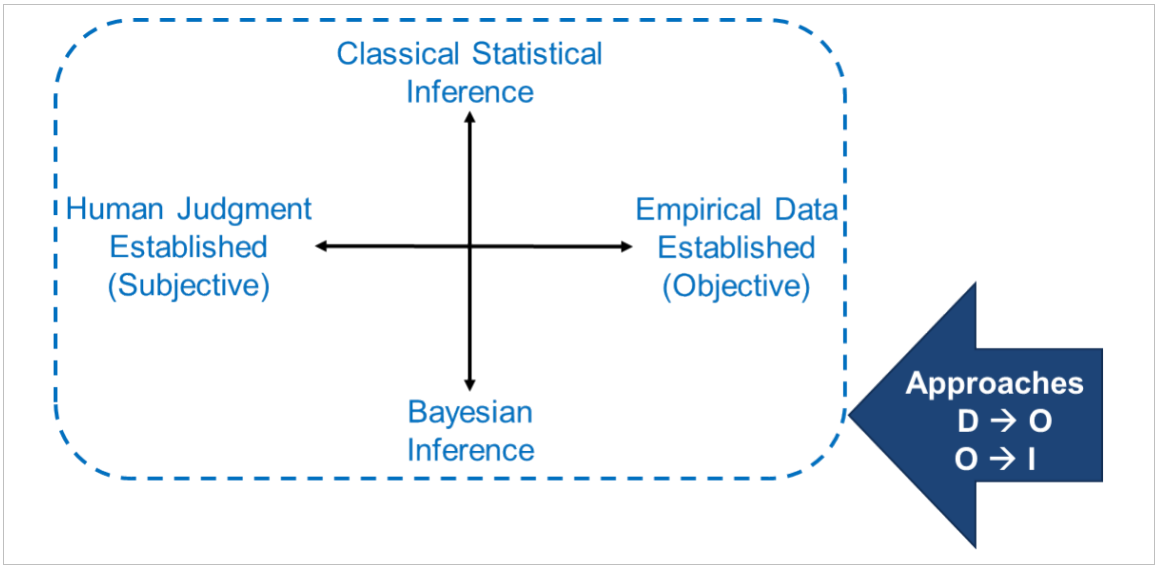

Figure 1: Landscape of Aggregate Indicator Measurement Methods

2.1 A Note on Ordinal-Scale Methods

Ordinal-scale methods have been historically popular to measure cybersecurity risk because of the complexity of that risk and the perception that many cybersecurity factors are impossible to measure more rigorously. These methods do not measure the factors in actual risk units. Rather, they measure factors on a multiple-point scale, such as a three-point scale (e.g., low/medium/high) or a five-point scale (e.g., low/low-medium/medium/medium-high/high). The important point is not how many points are on the scale, but rather that after being measured, the factors can only be ordered according to the group they are in, not ordered within the group they are in. In other words, we know that low factors are considered less than medium factors, and medium factors are less than high factors, but we can't assume an order of two factors in the same group.

Ordinal-scale methods are used extensively in cybersecurity risk management. Examples include the *Guide for Conducting Risk Assessments* [NIST 800-30 2012], the *Risk Rating Methodology* [OWASP 2020], and the *Common Vulnerability Scoring System* [CVSS 2020]. While these methods may be useful for ordering factors, they are not adequate for measuring the probability of attack, a component of risk associated with combining multiple indicators. The reason these methods aren't adequate for measuring probability is that any combination of indicators requires performing arithmetic operations on the measures of those indicators. Ordinal scales do not ensure the same interval between successive points, which is required for addition and subtraction; these mathematical operations require an interval scale measure. These methods also do not support multiplication or division since there is no notion of absolute zero in the scale; these mathematical operations require a *ratio-scale* method.

Researchers and practitioners reported these and other problems with using ordinal-scale measures for measuring cybersecurity risk [Hubbard 2014b, Cox 2008]. It is clear that ordinal- scale methods are inadequate as a basis for measuring the aggregation of indicators to estimate insider attack probability. Therefore, in the rest of this report, we focus on ratio-scale methods of measurement.

2.2 Types of Rigorous, Quantitative Methods Considered

In this section, we provide an overview of the four poles of the two-dimensional landscape shown in Figure 1.

Classical Statistical Inference: For our purposes, classical statistical inference uses classical statistical methods to develop measures of indicators that can justifiably be combined into measures of attack probability. An example is converting estimates of indicators into numbers on a ratio scale, as *normalized z-scores* do through normalization and centering estimates around a mean of zero [Dawes 1979]. More rigorous methods that involve statistical regression relate select values of x (in our case, an indicator) and an observed value of y (probability of insider attack). This general process is called *correlation*.

In the domain of insider threat, s*imple* regression relates a single indicator variable to attack probability, while *multiple* regression relates a set of indicator values to attack probability. The result of multiple *linear* regression is a best-fit line using the values of the indicator variables to "predict" the value of attack probability. This best-fit line is characterized using a set of weights (called coefficients) for

each of the indicator variables. More sophisticated *nonlinear* regression can find the curve that best fits the relationship of indicator values to insider attack probability.

Bayesian Inference: Bayesian methods use *conditional probabilities* to relate indicator probabilities to attack probability, instead of using classical statistical techniques . The logic of Bayesian methods allows calculating the probability of an insider attack given a set of indicators in terms of (1) the probability of the indicators given an insider attack, (2) the probability of the indicators given no insider attack, and (3) the (prior) probability of an insider attack before the indicators were observed. The indicators act as additional evidence that an insider attack will occur, which is used to update the (prior) probability of attack. These calculations (sometimes called Bayesian updating) are simplified somewhat if the indicator inputs are (or can reasonably be assumed to be) independent of each other. Therefore, the complexity of Bayesian measurement methods varies depending on how independent the inputs are assumed to be.

Human Judgment-Established Methods: People are often a good (and possibly the only) source of information for variable interrelationships. They have the ability to assess many complex and ambiguous situations that are impossible for most mechanical measurement methods; however, people are subject to a great many biases and fallacies that limit the accuracy of their estimation [Hubbard 2016, ch. 12]. Being an expert in a field does not usually reduce these limitations; in some cases, being an expert in the field seems to decrease the accuracy of their estimations.

Cybersecurity experts are no different; however, they still can provide valuable information about the relationship between indicators and attack probability as long as the source of the biases and fallacies can be addressed. Swets, Dawes, and Monahan developed techniques that yield the benefits of human judgment without being as hampered by its limitations [Swets 2000].

Empirical Data-Established Methods: Many analysis and measurement methods can accept empirical data as easily as it can accept expert judgment. By empirical data, we mean objective historical data that can help establish the relationship between indicators and attack probability. Such data is usually factual and is observed from actual system operations; thus, it is not subject to the biases and fallacies associated with subjective human estimates.

Because of the complexity of the problem, usually there is a fair amount of "noise" in the relationship between indicators and attack probability (e.g., false positives and missing indicators) that makes the relationship not straightforward despite the objectivity of the data. Using regression and Bayesian methods is helpful in clarifying the relationship in the face of this complexity.

2.3 Cost Matrix for Aggregate Indicator Measurement

Figure 2 shows the cost matrix for aggregate indicator measurement methods along the two dimensions discussed in the last section. Costs increase along the *human judgment established* dimension, the bottom row of the matrix, since more sophisticated analysis and estimation are required of the human expert. This analysis and estimation range from simple scoring of indicator relevance to estimating conditional probabilities of interdependencies. In addition, moving from *simple linear* to

correlation to *Bayesian inference* requires more samples across the range of interrelationships to ensure the sufficient accuracy of estimates.

	Simple Linear	Correlation	Bayesian
Empirical Data Established%	Updates z-score distributions with baseline and incident data, including incident prevalence	Uses baseline and incident data correlated with indicator prevalence to refine probability of insider attack	Uses Bayesian updating operating on the • network structure learned and refined through the analysis of baselines and incidents • priors, which are updated over time using data
Human Judgment Established	Normalized z-scores • expert estimates • weighted and additive estimates • static weighting	Correlation • best-fit line/curve that determines attack probability • coefficients that establish weights • design that reduces human inconsistency • static weighting	Bayesian Inference • expert estimates of baseline probabilities and conditional probabilities • certain interrelationships assumed to be conditionally independent • dynamic weighting

Figure 2: Indicator to Behavior Mapping Method Cost Matrix

Costs increase moving from human judgment to empirical data since data sources need to be identified and instrumented to provide data into the analytic and maintained as systems are updated and threats change. Finally, costs increase across the empirical data-driven dimension (the top row) as the following increases:

- sophistication of the analysis performed
- detail of the incident data needed regarding baseline and conditional probabilities
- the expertise needed to provide needed data
- the amount of support required for semi-automated analysis

3 Aggregate Indicator Measurement Methods

Figure 3 lists quantitative methods that are more rigorous and well-founded than the ordinal-scale methods described in the previous section. The rest of this section describes various methods and tools, and their substantial applications to cybersecurity (in general) or insider threat (specifically).

	General Methods	Insider Threat Tools
Classical Statistical Inference	• Normalized Z-Scores [Dawes 1979] • Lens Model [Brunswik 1955] • Rasch Model [Rasch 1961] • Log Odds Ratio (Stats) [Jaccard 2001]	• Forcepoint Behavioral Analytics [Forcepoint 2019] • Watchtower Behavioral Analytics [Arklay 2020]

Bayesian Inference	• Bayesian Networks [Pearl 1985, Neapolitan 1990]
	• Log Odds Ratio (Bayes) [Hubbard 2016, ch. 9]
	• Applications
	  • PNNL PACMAN [Greitzer 2010]
	  • ELICIT [Maloof 2007]
		• Haystax Security Analytics Platform [Haystax 2020]
		• Arcsight Interset Behavioral Analytics [Microfocus 2020]

Figure 3: Landscape of More Rigorous Quantitative Methods

3.1 From Simple Linear Models to Methods Based on Statistical Regression

The simplest human-judgment-driven methods are based on weighted scores of the major factors of importance in determining the variable of interest. In our case, it is the weighted score of indicators that measures insider attack probability. While the ordinal-scale measures described in the previous section are also weighted scores, at this more quantitative end of the domain, the methods do not use or reduce information to ordinal measures. That is, rather than relative and ambiguous measures, such as low-medium-high, these methods prefer whenever possible to use cardinal scales with concrete units, such as the frequency of downloads (in downloads per week) or the number of logins after normal working hours.

One such simple linear model is based on normalized z-scores [Dawes 1979]. After getting the judgment of the value of a particular factor by a group of human experts, a normalized distribution of the values is formed with an average of 0, and each value is translated into a number of standard deviations above or below that average. The values converted to deviations around the mean are called z-scores. The process of centering these scores around the mean of zero normalizes the factor value distribution so that it can be easily compared with other factors. This creates a ratio scale for the factors so that attack probability can be computed arithmetically.

Dawes' method performs better than the judgment of experts alone because the normalization of scores prevents one factor that has a wider range than another from taking on inordinate weight in the

calculation [Hubbard 2014a, ch. 12]. And in general, it has long been known that expert-based judgment is almost never as good as simple statistical models [Meehl 1954, Grove 2006, Swets 2000]. But while Dawes' method performs somewhat better than the unaided expert, it does not help with two broad classes of error in human judgment: random error and systematic error. For statistical models to perform significantly better than an unaided human, they must incorporate methods that reduce these error sources.

Systematic human error is a predictable and reproducible source of error due to flaws in human estimation capability, which manifests as regular deviation of the estimate from the true value. Research shows that people, including experts, are generally over confident in their estimation of their certainty about the probability of uncertain events. The more information they have, the more confident they become, even if the additional information bears no statistical relationship to what is being estimated. But Hubbard's research shows that people can be calibrated, much like mechanical instruments, to provide more accurate estimates [Hubbard 2014a, ch. 5]. This calibration involves estimation training and practice. Hubbard finds that after 20 years of human calibration experience, about 85% of trainees can reach calibration within a half-day of training.

Random human error, on the other hand, is an unpredictable and non-reproducible source of error in estimation because of uncertain changes in the environment. This type of error manifests as a random deviation of the estimate around the true value. The Lens model has been shown to remove human inconsistency—another name for random human error. Such inconsistency can account for 10% to 20% of the error in most expert estimates [Hubbard 2014a, ch. 12].

The Lens model takes as input an expert's judgments across a variety of situations including or not including various combinations of indicators. While expert judgment usually varies, even in identical situations, partly because of the biases and fallacies referenced earlier. The Lens model removes judgment inconsistencies by conducting statistical regression on the inputs to derive a best-fit mathematical formula that calculates the probability of attack associated with a set of indicators. This regression identifies the weights (coefficients) associated with each indicator.

Usually (linear) regression is used to identify a linear equation for the attack probability formula using the Lens method, but non-linear formulations may be advisable if a better fit is possible. Hubbard describes and exemplifies the use of conditional rules to perform piecewise linear regression [Hubbard 2014a, Exhibit 12.4]. Of course, the analyst must be careful not to overfit the data. Testing the formulation on a reserved portion of data is usually advisable.

Lens models are particularly well suited for removing inconsistencies of a single judge or a group of similar judges. Under these conditions, the Lens model consistently shows improvements over unaided expert opinion or over normalized z-scores reported by Dawes. In the case where there is a big (or unknown) difference between individual judges (e.g., some being very lenient or liberal in their estimations and others being very strict or conservative) and not all judges can judge every instance, the Rasch model provides a good alternative [Rasch 1961].

The Rasch model uses the log odds ratio method [Jaccard 2001] to even out the estimations of multiple judges so that the resulting estimations are as if one single consistent judge provided the

estimations. Using this model would provide a fair and even treatment of all insider threat indicators while considering their influence on attack probability.

3.2 Bayesian Methods

While methods based on statistical regression rely on correlation, which determines the weighting of the various indicators to calculate attack probability, Bayesian methods rely on conditional probabilities and Bayes Theorem to calculate attack probability. In its simplest form, and interpreted for the insider threat domain, Bayes Theorem can be expressed as the following:

$$P(InsiderAttack \mid Indicators) = P(InsiderAttack) * P(Indicators \mid InsiderAttack) / P(Indicators)$$

Where

$$P(Indicators) = P(InsiderAttack) * P(Indicators \mid InsiderAttack) + P(not\ InsiderAttack) * P(Indicators \mid not\ InsiderAttack)$$

The probability of an event (or outcome) x, which is expressed as P(x), can be thought of as a measure of the chance that x will occur, expressed as a percentage or a value between 0 and 1. The conditional probability is denoted P(x | y) and is expressed as "the probability of x given that y has occurred," where x and y are events or outcomes. While Bayes Theorem is not particularly intuitive, it is mathematically proven and very powerful.

In the above expression of Bayes Theorem, the probability of *InsiderAttack* is analogous to the calculation performed using statistical regression discussed in the last section. The event *InsiderAttack* can be conceptualized as any individual type of malicious insider compromise of interest, such as insider theft of intellectual property, sabotage, or fraud. The event(s) *Indicators* can be an individual indicator or any combination of indicators that might help predict a future or ongoing insider compromise.[3]

The (assumed) relationship between pairs of indicators helps the analyst understand the range of methods available for using conditional probabilities to calculate attack probability. The more individual indicators are assumed to be independent of one another (i.e., the value of one variable has no influence on the value of another), the simpler the computation of attack probability is. At one extreme, where variables are (assumed to be) completely independent of each other, the Log Odds Ratio method can be used to calculate insider attack probability based on the component conditional

[3] The expression P(*InsiderAttack*) is called the *prior probability* in relation to the conditional probability P(*InsiderAttack* | *Indicators*). The rest of the expression, i.e., P(*Indicators* | *InsiderAttack*) / P(*Indicators*), can be thought of as the probability of Indicators showing up given the insider compromise, which can be thought of as an adjustment to the prior. This adjustment is sometimes called the *posterior probability*.

probabilities [Hubbard 2016, ch. 9]. At the other extreme, a Conditional Probability Table (CPT) representing the relationship of every variable on every other variable can be determined.

While CPTs are the most conservative method of calculating conditional probabilities, they can be prohibitively expensive even for relatively simple problems, since the number of combinations explode with the number of variables represented. Since insider threat is a complex problem with no silver bullet indicators, many potential indicators are typically included in the detection mix.

This complexity all but rules out considering all potential interactions in a full-coverage CPT. Assuming away variable interactions, as in a naïve Bayes approach, is also not a good option since many indicators will have complex interactions. A good middle ground is to represent known significant interactions as conditional probabilities and assume independence in all other cases; Bayesian Belief Networks (BBNs) provide a rigorous, structured approach for achieving this middle ground.

BBNs are probabilistic, directed acyclic graphs where the nodes represent random variables, and edges represent conditional dependencies among random variables.[4] Conditional probabilities are what make BBNs Bayesian. Efficient algorithms exist to support the automatic application of Bayes Theorem through the network to the probability that a certain event will occur or a certain condition will exist. While the foundational work by Thomas Bayes was performed in the 1700s, it was not until Judea Pearl and Richard Neapolitan's writings in the 1980s that BBNs became a field of study [Pearl 1985, Pearl 1988, Neapolitan 1990]. BBNs have been applied to modeling knowledge in computational biology, medicine, bio-monitoring, document classification, information retrieval, semantic searching, image processing, data fusion, decision support systems, engineering, gaming, law, and risk analysis. BBNs can be developed using subjective human judgment or historical data. While a full discussion of applying BBNs as a basis for insider threat detection is beyond the scope of this report, we provide two prominent examples below.

3.2.1 Examples of BBNs

An example of developing BBNs for predicting insider threat strictly using human judgment is found in the work of Greitzer and Frincke [Greitzer 2010]. They considered 12 classes of indicators: disgruntlement, accepting criticism, anger management, disengagement, disregard for authority, performance, stress, confrontational behavior, personal issues, self-centeredness, lack of dependability, and absenteeism. As they note, not one of these indicators is likely to be a good predictor by itself because of their attendant high false-positive rates. Combinations of these indicators, however, can lead to good predictions of insider threat as they found out when they developed a BBN based in these indicators. The authors developed an initial BBN using two human resources experts to estimate indicator priors and conditional probabilities that establish the relationship among the indicators and insider attack probability. This initial work achieved a high level of correspondence between the model and

[4] BBNs can be very roughly thought of as a way of using a Lens model with conditional probabilities rather than correlation.

expert opinion (R^2=0.94), partly because the same experts used to develop the model were also used to test the model.

In follow-on work, Greitzer and Frincke more rigorously verified the model by conducting cross validation with 10 subject matter experts who rated the probability of insider compromise associated with each of 24 insider threat scenarios that represented various combinations of indicators [Greitzer 2012]. Using a round-robin approach, they used 9 of the experts' ratings to determine the conditional probabilities for the model and tested the instantiated model using the remaining expert's rating. The performance of the approach was very good (R2=0.598), which was commensurate with a parallel effort using linear regression.

Another example is the BBN approach developed by Maloof and Stevens. Their approach used both subjective human judgment and objective network data to determine priors and conditional probabilities [Maloof 2007]. They developed a system called *ELICIT* to detect need-to-know violations by malicious insiders. The indicators are described in a hierarchy [Caputo 2009, Figure 2] with *activity type* (e.g., searching) at the first level, *warning sign* (e.g., evasiveness, suspiciousness, and volumetric anomalies) at the second level, and *detector types* (such as the ones listed below) at the third level:

- Suspicious query timing is an example of evasive searching.
- Prohibited query item is an example of suspicious searching.
- High-query document retrieval is an example of a searching volumetric anomaly.

ELICIT uses 76 detectors that signal alerts for potential violations of an employee's need-to- know. Each detector is associated with a property that it is responsible for detecting. The ELICIT BBN assigns a threat score to users on the network based on the alerts that their activity generates. This BBN also has a three-level hierarchy:

1. The first level, the root node, is the Malicious Insider outcome. It specifies the prior probability that an insider is malicious: P(user is malicious).

2. The second level involves 76 nodes, one for each detector; it specifies the connection between user maliciousness and alerting via two probabilities:
   - P(user is malicious | user generated an alert for property p by the detector)
   - P(user is malicious | user did not generate an alert for property p by the detector)

3. The third level also involves 76 nodes; it specifies the connection between detector alerting and the property being alerted on via two probabilities:
   - P(detector will generate an alert for property p | p occurs)
   - P(detector will generate an alert for property p | p did not occur)

The probabilities at the first two levels were elicited from subject matter experts, while the probabilities at the third level were discovered from empirical data. This three-level tree- structured BBN specifies the potential complexity of handling all the combinations of the 76 detectors—and the 2^{76} possible combinations of alerts—while performing with good results. The authors tested the BBN using one month's worth of network traffic data with eight "red team" generated evaluation tests that

corresponded to real-world insider scenarios of the insider theft of information. Test results include a low false positive rate of 1.5% and an area under a receiver operating characteristic (ROC) curve of 0.92.

3.3 Discussion

There are tradeoffs associated with using classical statistical methods versus Bayesian methods with regards to aggregate indicator measurement of attack probability. Classical methods may be easier to apply, more intuitive in their application, and require less time and experience for human judges to supply necessary estimates. Establishing estimates for conditional probabilities and priors can be extremely challenging and important for the accuracy of aggregate measures [Green 2017]. Whether an organization uses human judges or empirical data largely depends on the existence of adequate historical data and the ability to mine that data along the required dimensions. For most organizations, a combination of human judgment and empirical data may be the best method that can be hoped for in the near term.

When human judgment is required, inconsistency of judgment (random error) can be eliminated using statistical regression as in the Lens method [Brunswik 1955]. Systematic error, for example due to overconfidence of the judge, can be reduced using calibration of the experts prior to estimation [Hubbard 2014a]. For human judgment analyzed using classical methods, Hubbard observes that the choice of analytic technique depends on the analysts' capability and resources:

> *It turns out that you can use weighted decision models at many different levels of complexity. If you feel confident in experimenting with non-linear methods, that's your best shot. If you can't do that but can handle linear regression, do that. If you don't feel comfortable using regression at all, stick with Dawes's equally weighted z-scores. Each method is an improvement on the simpler method, and all improve on unaided experts.* [Hubbard 2014a, chapter 12]

While Bayesian inference increases the complexity of the estimation task as well as the time and capability required of human judges, it does have some unique advantages over classical methods [Stuhl 2017a, Stuhl 2017b]. BBNs deal much better with missing values or extraneous variables than classical methods. BBNs use the whole distribution of variables' values rather than just individual correlations (e.g., one number summaries of how well variables align in a straight line), as in statistical regression. BBNs exploit all of the information available to derive aggregate measures. Extraneous variables in classical methods tend to excessively dampen the analyzed effect of the primary influencers. For classical methods to function well, approximate values must substitute for missing values.

In the insider threat domain, these are important considerations since it is far from clear exactly what variables should be included in the analysis. BBNs help conduct experiments in meaningful ways with a broad range of such variables, even if some turn out not to be that important in measuring insider attack probability. Finally, Hubbard reports significant improvement by moving from human judgment to empirical data when establishing a BBN, where that is possible and affordable [Hubbard 2014a, Exhibit 12.5].

4 Conclusion

In this paper, we characterized the primary methods used for measuring the probability of an insider attack by aggregating insider threat indicators, specifically focusing on those methods with the following characteristics:

- widely reported to be successfully used on the insider threat problem
- incorporated into prominent insider threat behavioral analytics tools
- consistent with the tools and capabilities observed at mature insider threat programs in government and industry

Previous CERT research focuses on the transform from observation (O) to indicator (I) and the associated cost matrix for methods supporting that transform. This report focuses on the transform from indicator (I) to behavior (B). Therefore, we focus on mapping indicators directly to the probability of insider attack.

The methods we describe range in sophistication from simple linear approaches to approaches using Bayesian inference to measure the aggregation of indicators. We discussed the tradeoffs and relative costs associated with these methods, providing a basis for organizations to decide how best to incorporate these methods into their insider threat behavioral analytics. We also identified several tools and applications that use these methods to further support organizations that want to improve their resources and capabilities. CERT researchers plan to release a subsequent report describing more sophisticated behavioral analytic techniques that can help organizations further build their foundational capability to more accurately identify and mitigate current and emerging insider threats.

References

[Arklay 2020]

Arklay. Watchtower Behavioral Analytics (WT-BA). 2020 [accessed]. https://www.arklay.net/products/

[Brunswik 1955]

Brunswik, Egon. Representative design and probabilistic theory in a functional psychology. *Psychological review*. Volume 62. Number 3. 1955. Page 193–217. https://psycnet.apa.org/record/1956-00035-001

[Caputo 2009]

Caputo, Deanna, Marcus Maloof, and Gregory Stephens. Detecting insider theft of trade secrets. *IEEE Security & Privacy*. Volume 7. Number 6. 2009. Pages 14–21. https://ieeexplore.ieee.org/document/5210090

[Cox 2008]
Anthony (Tony) Cox Jr, Louis. What's Wrong with Risk Matrices? *Risk Analysis: An International Journal* 28. Number 2. 2008. Pages 497–512. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1539-6924.2008.01030.x

[CVSS 2020]
Common Vulnerability Scoring System. Forum of Incident Response and Security Teams (FIRST). 2020 [accessed]. https://www.first.org/cvss

[Dawes 1974]
Dawes, Robyn M. & Corrigan, Bernard. Linear models in decision making. *Psychological Bulletin*. Volume 81. Number 2. 1974. Pages 95–106. https://psycnet.apa.org/record/1975-22200-001

[Dawes 1979]
Dawes, Robyn M. The robust beauty of improper linear models in decision making. *American Psychologist.* Volume 34. Number 7. 1979. Pages 571–582. https://psycnet.apa.org/record/1979-30170-001

[Forcepoint 2019]
Forcepoint. *Forcepoint Behavioral Analytics User Manual V3.3*. Forcepoint. 2019. https://www.websense.com/content/support/library/ueba/v32/user_manual/user_manual.pdf

[Freund 2014]
Freund, Jack & Jones, Jack. *Measuring and Managing Information Risk: A FAIR Approach*. Butterworth-Heinemann. 2014. https://www.elsevier.com/books/measuring-and-managing- information-risk/freund/978-0-12-420231-3

[Green 2017]
Green, Michael. The Truth About Bayesian Priors and Overfitting. *KDNuggets*. 2017. https://www.kdnuggets.com/2017/07/truth-about-bayesian-priors-overfitting.html

[Greitzer 2010]
Greitzer, Frank L. & Frincke, Deborah A. Combining Traditional Cyber Security Audit Data with Psychosocial Data: Towards Predictive Modeling for Insider Threat Mitigation. In *Insider Threats in Cyber Security*. Springer. Pages 85–113. 2010. https://link.springer.com/chapter/10.1007%2F978-1-4419-7133-3_5

[Greitzer 2012]
Greitzer, Frank L.; Kangas, Lars J.; Noonan, Christine F.; Dalton, Angela C.; & Hohimer, Ryan E. Identifying At-Risk Employees: Modeling Psychosocial Precursors of Potential Insider Threats. Pages 2392-2401. In *2012 45th Hawaii International Conference on System Sciences*. IEEE. 2012. https://ieeexplore.ieee.org/document/6149305

[Greitzer 2018]
Greitzer, Frank; Purl, Justin; Leong, Yung Mei; & Becker, DE Sunny. SOFIT: Sociotechnical and Organizational Factors for Insider Threat. Pages 197-206. In *2018 IEEE Security and Privacy Workshops (SPW)*. 2018. https://www.researchgate.net/deref/http%3A%2F%2Fdx.doi.org%2F10.1109%2FSPW.2018.0003

[Grove 2005]
Grove, William M. Clinical versus statistical prediction: The contribution of Paul E. Meehl. *Journal of Clinical Psychology*. Volume 61. Number 10. 2005. Pages 1,233–1,243. https://doi.org/10.1002/jclp.20179

[Grove 2006]
Grove, William M. & Lloyd, Martin. Meehl's contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology*. Volume 115. Number 2. 2006. Pages 192–194. https://psycnet.apa.org/record/2006-06737-002

[Harker 1987]
Harker, Patrick T. & Vargas, Luis G. The Theory of Ratio Scale Estimation: Saaty's Analytic Hierarchy Process. *Management Science*. Volume 33. Number 11. 1987. Pages 1,383–1,403. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.33.11.1383

[Haystax 2020]
Haystax Security Analytics Platform. 2020 [accessed]. https://haystax.com/platform/

[Holt 1958]
Holt, Robert R. Clinical and statistical prediction: A reformulation and some new data. *The Journal of Abnormal and Social Psychology*. Volume 56. Number 1. 1958. Pages 1–12. https://psycnet.apa.org/doi/10.1037/h0041045

[Hubbard 2012]
Hubbard, Douglas W. *The Failure of Risk Management: Why It's Broken and How to Fix It*. Wiley Publishing. 2012. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119198536

[Hubbard 2014a]
Hubbard, Douglas W. *How to Measure Anything: Finding the Value of Intangibles in Business*. Wiley Publishing. 2014. https://www.wiley.com/en- us/How+to+Measure+Anything%3A+Finding+the+Value+of+Intangibles+in+Business%2C+3rd+Edition-p-9781118539279

[Hubbard 2014b]
Hubbard, Douglas W. *The failure of risk management: Why it's broken and how to fix it*. John Wiley & Sons. 2014.

[Hubbard 2016]
Hubbard, Douglas W. & Seiersen, Richard. *How to Measure Anything in Cybersecurity Risk*. Wiley Publishing. 2016. https://www.wiley.com/en- us/How+to+Measure+Anything+in+Cybersecurity+Risk-p-9781119085294

[Jaccard 2001]
Jaccard, J. *Interaction Effects in Logistic Regression*. Issue 135. SAGE Publications. 2001. Rotella, J. (n.d.) "Probability, log-odds, and odds." http://www.montana.edu/rotella/documents/502/Prob_odds_log-odds.pdf

[Maloof 2007]
Maloof, Marcus A. & Stephens, Gregory D. ELICITELICIT: A System for Detecting Insiders Who Violate Need-to-Know. Pages 146-166. In *International Workshop on Recent Advances in Intrusion Detection*. Springer. 2007. https://doi.org/10.1007/978-3-540-74320-0_8

[Meehl 1954]
Meehl, Paul E. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. 1954. https://psycnet.apa.org/record/2006-21565-000

[Microfocus 2020]
Microfocus. *What Makes Interset Different (*Parts 1-3). 2020 [accessed].
https://www.microfocus.com/media/white-paper/what-makes-interset-different-part-1-wp.pdf
https://www.microfocus.com/media/white-paper/what-makes-interset-different-part-2-wp.pdf
https://www.microfocus.com/media/white-paper/what-makes-interset-different-part-3-wp.pdf

[Neapolitan 1990]
Neapolitan, R. E. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley- Interscience. 1990.

[NIST 2012]
National Institute of Standards and Technology (NIST). *Special Publication (SP) 800-30, Revision 1: Guide for Conducting Risk Assessments*. NIST Computer Security Resource Center. 2012.

[OWASP 2020]
Open Web Application Security Project (OWASP). *OWASP Risk Rating Methodology*. OWASP. 2020 [accessed]. https://owasp.org/www-community/OWASP_Risk_Rating_Methodology

[Pearl 1988]
Pearl, Judea. *Probabilistic Reasoning in Intelligent Systems*. Elsevier. 1988. https://www.elsevier.com/books/probabilistic-reasoning-in-intelligent-systems/pearl/978-0-08- 051489-5

[Pearl 1985]
Pearl, J. Bayesian Networks; A Model of Self-Activated Memory for Evidential Reasoning. Pages 329-334. In *Proceedings of the 7th Conference of the Cognitive Science Society*. 1985.

[Perez 2006]
Pérez, Joaquín; Jimeno, José L.; & Mokotoff, Ethel. Another Potential Shortcoming of AHP. *Top*. Number 14. Volume 1. 2006. Pages 99–111. https://doi.org/10.1007/BF02579004

[Rasch 1961]
Rasch, Georg. On General Laws and the Meaning of Measurement in Psychology. Pages 321–333. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Volume 4. 1961.

[Rich 2019]
Rich, Charley, Pankaj Prasad, and Sanjit Ganguli, Gartner Market Guide for AIOps Platforms, 2019. https://www.gartner.com/en/documents/3971186/market-guide-for-aiops-platforms

[Saaty 1987]
Saaty, Roseanna W. The Analytic Hierarchy Process—What It Is and How It Is Used. *Mathematical Modelling*. Volume 9. Number 3-5. 1987. Pages 161–176.

[Saaty 1980]
Saaty, Roseanna W. *The Analytic Hierarchy Process*. McGraw Hill. 1980.

[Savage 2006]
Savage, Leonard J. *The Foundations of Statistics*. Courier Corporation. 1972.

[Slovic 1971]
Slovic, Paul & Lichtenstein, Sarah. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*. Volume 6. Number 6. 1971. Pages 649–744.
https://www.sciencedirect.com/science/article/pii/003050737190033X

[Sticha 2016]
Sticha, Paul J. & Axelrad, Elise T. Using dynamic models to support inferences of insider threat risk. *Computational and Mathematical Organization Theory*. Volume 22. Number 3. 2016. Pages 350–381. https://link.springer.com/article/10.1007/s10588-016-9209-1

[Struhl 2017b]
Struhl, Steven. How Bayesian Networks Are Superior in Understanding Effects of Variables. *KDNuggets*. 2017. https://www.kdnuggets.com/2017/11/bayesian-networks-understanding-effects-variables.html

[Struhl 2017a]
Struhl, Steven. The amazing predictive power of conditional probability in Bayes Nets. *KDNuggets.* 2017. https://www.kdnuggets.com/2017/11/amazing-predictive-power-conditional-probability-bayes-nets.html

[Swets 2000]
Swets, John A.; Dawes, Robyn M.; & Monahan, John. Psychological Science Can Improve Diagnostic Decisions. *Psychological Science in the Public Interest*. Volume 1. Number 1. 2000. Pages 1–26. https://journals.sagepub.com/doi/10.1111/1529-1006.001

[Tabish 2016]
Rashid, Tabish; Agrafiotis, Ioannis; & Nurse, Jason RC. A New Take on Detecting Insider Threats: Exploring the Use of Hidden Markov Models. Pages 47-56. In *Proceedings of the 8th ACM CCS International Workshop on Managing Insider Security Threats*. 2016. https://www.cs.ox.ac.uk/files/8364/MIST2016-RAN-AuthorFinal.pdf

[Yuan 2018]
Yuan, Fangfang; Cao, Yanan; Shang. Yanmin; Liu, Yanbing; Tan, Jianlong; & Fang, Binxing. Insider Threat Detection with Deep Neural Network. Pages 43-54. In *International Conference on Computational Science*. 2018. https://www.iccs- meeting.org/archive/iccs2018/papers/108600047.pdf

Legal Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM24-0725

Contact Us

Software Engineering Institute
4500 Fifth Avenue, Pittsburgh, PA 15213-2612

Phone: 412/268.5800 | 888.201.4479
Web: www.sei.cmu.edu
Email: info@sei.cmu.edu