VivoSecurity
Automated Risk Quantification

Feb 17, 2017

Robert de V. Frierson, Secretary
Board of Governors of the Federal Reserve System
20th Street and Constitution Avenue NW
Washington, DC 20551

Legislative and Regulatory Activities Division
Office of the Comptroller of the Currency
400 7th St. SW
Suite 3E-218, Mail Stop 9W-11
Washington, DC 20219

Robert E. Feldman, Executive Secretary
Federal Deposit Insurance Corporation
550 17th St. NW
Washington, DC 20429

**Re: Advance Notice of Proposed Rulemaking on *Enhanced Cyber Risk Management Standards*; Federal Reserve Docket No. R-1550, RIN 7100-AE-61; OCC Docket ID OCC-2016-0016; FDIC RIN 3064- AE45**

Dear Ladies and Gentlemen,

This comment is submitted on behalf of VivoSecurity Inc. for consideration regarding the *Enhanced Cyber Risk Management Standards* proposed in the Advance Notice of Proposed Rulemaking (ANPR) of October 19, 2016. VivoSecurity is a Silicon Valley startup company focused on quantifying and profiling cyber risk in dollar terms. Observations and comments herein are based on our experience working with financial institutions and other clients.

In section VII ("Approach to Quantifying Cyber Risk") of the ANPR, the three agencies call for a "consistent repeatable methodology" for assessing and measuring cyber risk. We support this goal, noting that cybersecurity has reached a level of impact that warrants the rigorous management discipline and oversight that FIs routinely apply to more established domains of business and operational risk. Our experience shows that cyber risk can be objectively quantified using empirical data and careful statistical analysis. In this comment we will argue that such an approach is both realistic and necessary in the context of regulatory supervision of banking entities' operational risk.

## Antecedents and Rationale

Because cyber is a new area, perceived as highly volatile, it is often assumed that rigorous quantification is infeasible. Indeed, the conventional "risk matrix" approach has been demonstrated to be ineffective[1], often leading to worse risk mitigation decisions than would have been made in the absence of any measurement procedure. Risk matrices using ordinal scales (e.g. "red, yellow, green") suffer from several weaknesses. Expert ratings of risk factors are usually subjective, undefined, and ungrounded; often little more than guesses[2]. Risk factor phrasing may be imprecise and inconsistently interpreted, leading to disparities in assessments. Ratings are prone to clustering and range compression, which can sharply reduce accuracy especially when multiple ordinal categories are combined onto a matrix. In general, with ordinal scales it is difficult or impossible to combine individual factor scores into a meaningful consolidated assessment either for business risk management or simply for prioritization and resourcing of mitigation efforts.

Newer approaches eliminate the risk matrices and ordinal scales in favor of quantitative estimates of expected frequency and magnitude of loss. Hubbard[3] has demonstrated the efficacy of calibration and statistical techniques to reduce subjectivity and increase accuracy in experts' estimates of ranges of values of underlying risk factors. He and other practitioners use such tools to forecast risk in various domains. However the credibility and utility of these techniques for cyber risk is limited by two factors:

1) Forecasts of likelihood and severity are calculated using random-attack simulations (typically Monte Carlo) based on theoretical models of breach dynamics. The complexity of cyber and constantly changing interactions among threats, vulnerabilities, controls and assets prevent any effective validation of the models. Our concern here is that simulations based on theory too often give a false sense of validity to theory that has no empirical foundation.

2) Model inputs and parameters are still subjective, based on human estimates. Even calibrated and controlled, these estimates suffer from lack of contextual normalization.

---

[1] We make no attempt to survey the considerable literature on this point, or on quantitative risk assessment in general. However we would like to highlight the books of Douglas W Hubbard, in particular the 2016 book by Douglas Hubbard and Richard Seiersen, *How to Measure Anything in Cybersecurity Risk* (John Wiley & Sons). The authors cite peer-reviewed research studies on effectiveness of various risk assessment methodologies and visualization tools.

[2] An additional reference is Daniel Kahneman, 2002 Nobel laureate in Economics for research showing, among other things, that simple heuristics developed by an intelligent lay person generally outperform experts, unless the experts routinely receive high quality feedback in a timely fashion. Such feedback is difficult to obtain in cyber security. See for example his 2013 book, *Thinking, Fast and Slow* (Farrar, Straus and Giroux).

[3] See earlier footnote

The latter is best seen by example. Consider a cybersecurity expert estimating upper and lower bounds at some level of confidence for various elements in the frequency and/or severity of a loss caused by a particular control deficiency (perhaps a missing patch set) relative to a given type of attack that compromises a specific set of assets. Now suppose that the expert has access to data, either historical or analytical, to inform the estimates. For example he or she might have prior experience with the costs of forensic investigation or legal services related to such a breach, or knowledge of the per-customer costs involved in notification or credit monitoring services. With such data the expert can make a reasonably accurate forecast of cost factors for the model under construction. Without it, the expert's estimates are essentially guesses. While calibrated expert estimates are arguably better than no information at all, they reflect assumptions and opinions of the expert (or collective assumptions of a group of experts, likely biased by inaccurate portrayals of cyber risk in industry and mainstream media) and in practice are often strikingly inaccurate. Worse, there is no way to model or measure either the accuracy of the expert estimates or the error as represented by the expert's supposed confidence intervals.

Hence, historical or analytical data are key to the accuracy and credibility of the risk measurement. We propose then to apply that data directly, using a simpler and more deterministic modeling methodology to forecast risk. Availability of data is a crucial enabler, a topic we will address below.

Modeling cyber risk in financial institutions

We have used a variety of regression-based empirical modeling techniques to forecast frequency and cost of cyber incidents for clients in financial services and other regulated industries. There are two main advantages of regression models, compared with simulation techniques. First, tools for multiple regression (including ensemble and machine learning methods) allow analysis of the role and impact of individual characteristics of a cyber incident. Breaches have attributes relating to the organization affected, the nature of the attack or incident, the data or other assets compromised or damaged, etc. Multiple regression is used to identify and quantify the relationship between these potential *predictor* variables and the frequency and/or cost of cyber incidents.

The result is an evidence-based approach to measuring and understanding risk. In contrast to simulation style modeling, in which the analyst makes a series of assumptions about the factors and causes of incidents and then relies on subjective estimation to set coefficients for those factors, the regression methodology simply follows the data wherever they lead. In fact, we find that the data often lead in surprising directions, offering unexpected insights into factors behind cyber vulnerabilities or attack vectors (examples below).

Understanding the factors and their weighting allows both for accurate quantification of the risk and for cost-effective risk mitigation and management strategies.

The second advantage of regression is characterization of error. Combined with careful filtering and pre-processing of model input data, we obtain statistically grounded estimates of the variance (or confidence intervals) of risk predictions made using the model. Confidence intervals derived from empirical data, using statistical mathematics, increase the credibility and accuracy of the risk forecast – including its imprecision.

Statistical error metrics take on special significance for financial institutions because of the scrutiny in recent years on model risk. Regression models are subject to measurement, testing, and validation with respect to solidity, reliability, and suitability for risk management. Models can easily be developed and tested in conformance with the FRB/OCC SR 11-7 Model Risk Management Guidelines and the model validation procedures of individual financial companies.

Availability of data for modeling

It is generally assumed in the cybersecurity world that little or no suitable data are available for construction of quantitative risk models – first, because affected organizations do not share information about breaches and especially about costs incurred, and second because cyber breaches are skyrocketing chaotically and cannot be predicted.

Both of these assumptions are myths. We have found substantial data on historical cyber incidents from a variety of sources. Most states, and federal regulatory agencies in some industries, require reporting of cyber breaches and in many cases make the data available. Further information is available in SEC filings, news accounts of legal settlements and regulatory actions, industry reports, and other sources. While the data are indeed incomplete, they provide ample basis for training sets for statistical models.

Note that the available breach data lend themselves to our top-down regression-based approach to modeling. We are not seeking empirical or historical data on low-level interactions among specific threats, attack vectors, or controls. Such data are indeed scarce, which is the reason simulation-based approaches rely on non-empirical expert estimation.

We and others have also found that the tabloid-driven perception of cyber breaches spiraling out of control is inaccurate (and a source of bias in methods based on expert opinion, as indicated above). The rate of breaches in recent years, especially when correlated with selected predictor variables as discussed above, is constant or even in decline in some categories. Using regression tools we find patterns in the cyber breach data

that are measurably stable and consistent over time, and thus have predictive value with regard to future events.

Furthermore cyber events and data constitute a popular topic of research in both companies and universities. Many of the research studies collect and analyze information on effects of specific cybersecurity controls and countermeasures, and the resulting data can increase the granularity and operational relevance of quantitative risk models. We provide two brief examples here:

- A study by Symantec showed a strong correlation between an individual's chance of being targeted in a spear phishing attack and the number of connections the person has on LinkedIn – surprisingly, an *inverse* correlation (people with very few connections are more often targeted). LinkedIn connections were a stronger predictor than organizational level or role.
- Our own research, based on vendor data, shows in quantitative terms the influence of operating system release on a computer's vulnerability to malware. Going beyond the simple observation that more recent releases are more secure, we can use OS release along with other information to forecast frequency of occurrence of malware attacks.

These and similar examples show the potential for a quantitative risk profile, generated using historical data and empirical observations, in guiding effective management, mitigation, and governance of cyber risk.

Conclusion

In this comment we have argued that cyber risk can be being measured in dollar-quantified terms, using mature statistical modeling techniques based on purely empirical data. This is likely the only approach to risk quantification that can provide the consistency and rigor sought in the ANPR and can also satisfy the model risk scrutiny appropriate to the financial sector. A dollar-quantified statistical risk profile can guide operational cyber risk management and mitigation activities, inform cyber insurance coverage requirements, and feed into the enterprise-wide assessment of reserves and coverage needed for operational risk.

We recommend that the financial regulatory agencies adopt a mandate for statistical cyber risk quantification grounded in empirical data, as part of the adoption of the Enhanced Cyber Risk Management Standards for medium- and large-scale financial institutions. Such a mandate will foster a culture of rigorous risk management in the banks, their business partners and vendors, and in the risk management industry. It will lead to increased

funding and interest in research studies to augment the base of empirical data and hence grow the granularity, usefulness, accuracy, and reliability of statistical risk models. This virtuous circle will bring cybersecurity into a new generation of business maturity through data-grounded risk management commensurate with the importance of this new risk area to the stability of the financial sector.

We at VivoSecurity Inc. appreciate the opportunity to provide input to the important efforts underway to strengthen cybersecurity and cyber governance in this industry.

Sincerely,

James Lipkis
COO and General Manager, VivoSecurity Inc.
jiml@vivosecurity.com
408-219-0450