QOMPLX:

Honesty About Effective Modeling Cyber Risk for Insurance Companies

By Jason Crabtree and Alastair Speare-Cole

Modelers are sometimes promising a blow-by-blow account of an event when they only have a few snapshots of what happened.

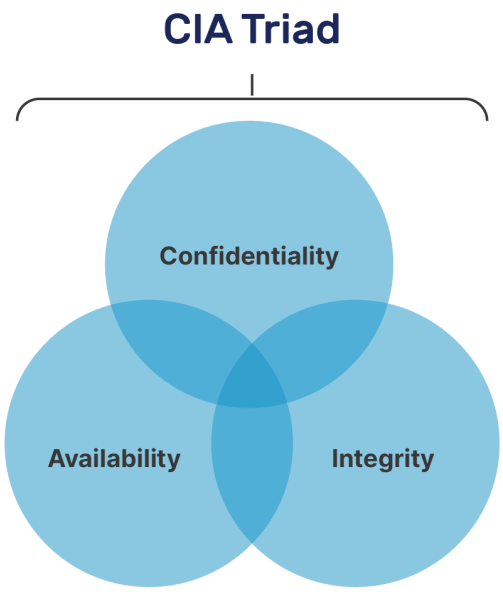
As cybersecurity plays a larger and larger role in daily lives and the global economy, there is growing importance in our collective quest to accurately understand and price cyber-related risks. Doing this correctly, transparently, and reliably is an important part of engendering better behavior in both public and private institutions as a function of governance. However, in this nascent industry, cyber risk modeling companies and some carriers are seeking to profit off the opportunities for growth in a new class of business without adequate consideration of the tail risk. In addition to questions about specific policy wordings (e.g. attribution-related clauses which are inappropriate at best), many seem to be attempting to drown clients and regulators in a tidal wave of "information" that poorly correlates to whether or not a company will be breached. External scan data is necessary but not sufficient for modeling cyber risk in terms of either frequency or severity.

All too often, current data brokers and modelers make bold assertions about their ability to correctly model the core aspects of cyber-related risks with models that inadequately consider the context necessary to be predictive. The novelty and uniqueness of modeling cyber-related risks and the difficulty of correctly linking measures of exposure to the propensity to incur losses should mean that all modeling claims be met with substantial caution. Any honest modeling firm, broker or (re)insurer should be candid with their clients about modeling approaches and associated limitations. There are three key features unique to cyber that make both accuracy and precision difficult for even deterministic models - further amplified when attempting to create stochastic models: (1) access to the data necessary to build a holistic model, (2) the heterogeneity of the ultimate risk pool under consideration, and (3) proper data collection, aggregation and sampling.

Data Accessibility

First, access to the data necessary to build successful and sufficiently complete models is a major current challenge. Although a substantial amount of data is collected by insureds, it is held in fragments with no incentive to share it. Further, historical loss events and claims handling processes do not really link the pre-event health data or telemetry with mid-event or post-event data. In other words, modelers are sometimes promising a blow-by-blow account of an event when they only have a few snapshots of what happened.

There is no obligation to disclose the specifics of breaches that could be used to improve modeling. Or even more substantially, attacks not known to have caused a data loss event but which may have caused integrity or availability issues. This has also meant that most models being employed are limited exclusively to Confidentiality concerns and ignore Integrity and Availability concerns common to the "CIA" triad typically referenced in information security.



This challenge is compounded by the current approach to network and organization characterization for both third-party risk and health characterization. Companies today are far too focused on the limited external indicators of organizational security, even though modern network architectures and segmentation severely limit visibility from external sources and therefore render many of these metrics dubious at best. Internal metrics about network state, health, and configuration are much more informative and important when considering the organizational security posture and its correlation with eventual losses or lack thereof (at least outside of the most grossly negligent errors).

Some modelers think that, even with limited data, they can establish correlations that inform us as to the underlying causes of loss. However, even when using the most methodical process and logical criteria, limited data means that models can only reveal relative risk. While it is a helpful metric, there is a substantial difference between relative and absolute risk. Additionally, models are currently being fit to historical data that may have little to do with the evolving tactics, techniques, and procedures (TTP) of both state and non-state threat actors in cyberspace. Traditional actuarial techniques focus on extrapolation of historical experience within a given peril into the future based on consistent relationships between various parameters. In cyber, the dynamic between attacker and defender mixed with a devastating rate of technological change. For example, machine learning (ML) for anti-virus and malware detection was initially quite promising, but binary executable transforms and other techniques already make it possible to create unique malware that can easily evade even the latest ML systems. Similarly, memory-related attacks which were common several years ago are less common today as credential-based exploits and tooling have massively increased, enabling stealthier breaches which leverage lateral movement techniques to avoid detection for longer periods of time. In other words, the ongoing changes across people, process, technology, and data sources (on the side of both attackers and defenders) means that we should continue to look at all the data available while

Cyber's collection of heterogeneous risks leads to a constantly shifting universe.

recognizing that some retrospective data may not maintain value or meaning as the techniques and capabilities of participants continue to evolve. As new threats and vulnerabilities continue to arise, retrospectively dominated forecasting and modeling techniques are not likely appropriate. A focus on simulation-based techniques is best in order to explore a broader range of futures rather than extrapolating from past events.

Risk Pool Heterogeneity

The second problem in modeling cyber risk is more profound: the risk pool is extremely diverse. Understanding financial risk requires an understanding of a heterogeneous group of threat actors (including strategically motivated nation-states and non-state actors with their own economic motives), the companies/assets themselves (with different levels of investment and culture), and ultimately the business impact (stemming from substantially different process exposure to the underlying IT systems and data). Typically, this means we develop a comprehensive view of the assets of interest, a vulnerability model for those assets, a hazard model, and then a financial model which can be assessed across each of the many different scenarios to be evaluated.

The practical dynamics of cyber risk as an adversarial domain where a victim can be either a strategic or opportunistic target actually undermines a typical insurance industry approach of treating clients and losses as a homogeneous risk pool. Additionally, unlike other domains of risk assessment such as natural catastrophe, geography is often not a useful factor when looking at potential risk accumulation since common technologies, services, or assets do not cluster or have impact limited to traditional geographic boundaries. In a largely virtual world such boundaries have little meaning. Cyber's collection of heterogeneous risks leads to a constantly shifting universe, making accurate assessment more difficult and challenging than geography- and sector-based accumulation management approaches. Cyber requires more dynamic and diverse factors to be accounted for to address accumulation risks.

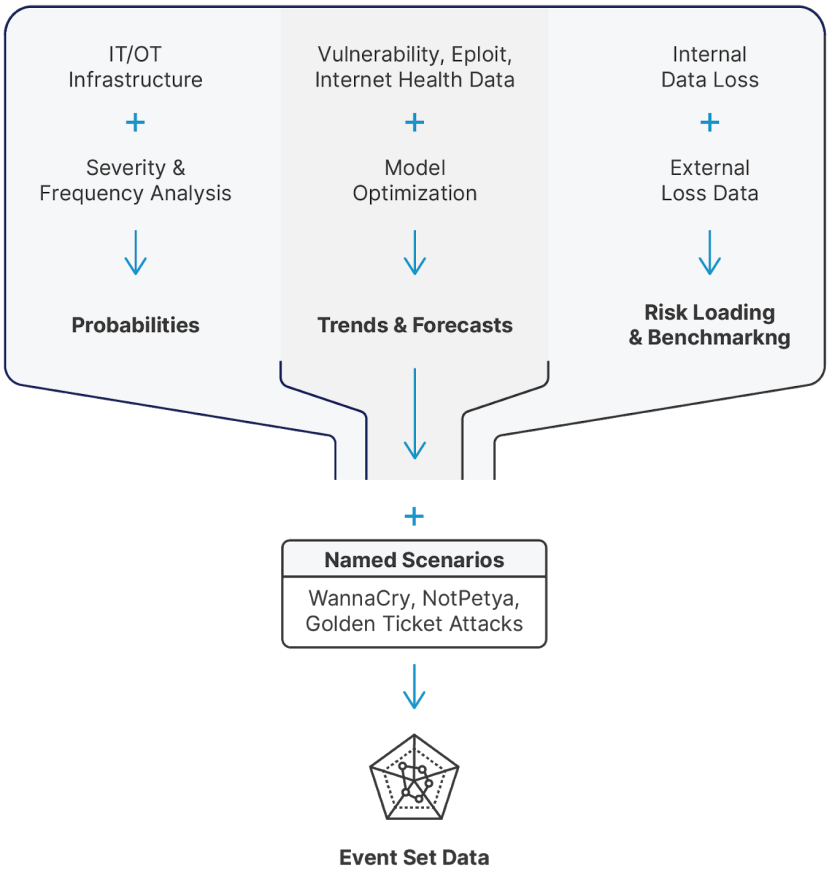
Another common claim from modelers is that they have completed the development of full stochastic models. Such claims should currently be taken with a grain of salt. Purely stochastic modeling approaches are completely inappropriate for cyber. While the ability to sample from a range of distributions re: hazard, vulnerability, asset composition, and financial loss is a step towards a sustainable modeling regime - the continued quest for precision without accuracy remains foolish.

Cybersecurity models must address the inconvenient truth that adversaries can learn, and the models must reflect that. Traditional stochastic modeling techniques from natural perils are inappropriate for environments like cybersecurity and terrorism where adversaries learn. Events are not random samples of a distribution. There are ways of more intelligently sampling - but hurricanes and storms are not sentient beings with specific intent or the ability to transfer knowledge about tactics, techniques, procedures or even impacts from others exploits.

Data Collection, Aggregation, and Sampling

The third key factor in cyber risk modeling is understanding that data collection, aggregation, and sampling is complex for cyber. The insurance industry is familiar with the construction of event sets and the ultimate event loss tables (used to describe the frequency and severity) which can be analyzed across factors such as geography. Standard approaches work to ensure that exposures and expected losses do not threaten the overall portfolio and can be critical to decision-making for underwriters, risk managers, and regulators alike. In the traditional property world, the perils are relatively static and follow reasonably consistent patterns where expected loss severity correlates strongly to property values and the types of disasters capable of reaching that locale. As tragic as natural disasters are, a hurricane never attacked an individual or community. A wildfire has not been constructed and directed with the intent to inflict maximum damage. In cyber security, the adversary is sentient, using intentionally unpredictable methods to achieve specific goals based on diverse incentives that may include financial gain or service disruption.

Event Set Generation



These are just some of the reasons why cyber presents a series of tough challenges to the insurance industry. The hazard model can change drastically due to geopolitical actors sparring. The vulnerability model must account for everything from

Tremendous strides have been made in the early days of cyber risk assessment. But this will be a long road to travel.

Heartbleed-like bugs that can be quickly patched to Spectre-like issues that might be present for a decade. This means that risk accumulation can appear or change in size almost instantly. Risk capital will not be resized annually in such a market - at least not at high levels of market penetration. For accumulations resulting in unsuitable tail risk, it must be possible to describe, measure, codify, and ultimately transfer the risk to address the challenges posed by both affirmative cover and non-affirmative or cover which represents a substantial unaccounted exposure for the industry.

Modeling claims for cyber should be viewed with caution and thought about in an adversarial fashion. Simple scans and promises of stochastic model availability offer a false sense of security to underwriters, risk managers, and regulators. In addition to a more honest dialogue about modeling challenges, a true discussion regarding data and telematics is required.

Before making a decision, consider the following five areas of interest and concern that should be top-of-mind when considering any modelling solution. First, models that do not offer open architecture capable of incorporating new datasets, if or when available, should be rejected. Second, any model not capable of being extended or considering confidentiality, integrity and availability of data/services should be rejected. Third, exact calculation of exposures must be made visible, to enable both affirmative and non-affirmative cover to be modeled. Fourth, the frequency of model calibration (along with calibration methods) and a comprehensive list of data sets used to calibrate should be provided to aid in understanding model risk associated with data being out of step with the present or likely future. Fifth, when a modeler claims that a model can relate the history of a company's cyber events to its technology, any insurer should insist on understanding the metrics and the basis for that claim - external scans and user-provided data alone should be viewed as highly suspect.

Tremendous strides have been made in the early days of cyber risk assessment. But this will be a long road to travel. The industry has only just stepped out of the gate and onto the road, and not all who offer direction are acting in the best interests of insureds or (re)insurers. A scientific approach that uses retrospective and generative modeling techniques will be of the utmost importance to building a sustainable and data-driven basis for cyber and operational risk transfer.