Next-Gen Scoring

A New Standard in Fair and Transparent Cybersecurity Ratings

SecurityScorecard

Overview – Scoring Fundamentals of Accuracy, Fairness, and Transparency

In June 2017, SecurityScorecard committed to the U.S. Chamber of Commerce's brief for delivering transparent, fair, and accurate security ratings. Security ratings can provide insights enabling better security when the companies creating the scores explain the algorithms so that businesses can understand how to use them meaningfully. In SecurityScorecard's continued and proactive transparency efforts, we are excited to introduce our Next-Gen Scoring. This whitepaper explains how Next-Gen scoring brings increased accuracy and how the methodology works.

Motivation

A unique challenge in providing fair and accurate ratings for organizational cybersecurity is properly accounting for the wide dynamic range in organizational size.
Smaller entities, such as "MomAndPop.com" bearing a small digital footprint with just a single or a few IPs, will inevitably have fewer findings and correspondingly fewer security flaws compared to large enterprises operating over as many as hundreds of millions of IPs. Conversely, larger entities will nearly always have more security defects than smaller entities and would receive worse security scores if no correction for entity size were made.
Conversations with customers, security experts, and data scientists led our product teams to introduce scoring methods that not only stay true to our principles of accuracy, transparency, and fairness, but also provide new ways to reduce ecosystem risk. Fairness means accounting for the varied risks associated with different organizations within an ecosystem and finding ways to compare various digital footprints.

To address this unmet need, the Data Science team at SecurityScorecard developed a principled scoring methodology based on a robust, statistical framework that ensures fair scores regardless of organization size. Dubbed Next-Gen Scoring, the new scoring system also enables users to better understand their scores and to develop a quantitative plan with recommendations for which specific issues to address in order to remediate their cybersecurity scores by a user-specified amount.

How Next-Gen Scoring Enables Better Security

The introduction of more sophisticated statistical models that underpin Next-Gen scoring will allow for novel insights that enable organizations to prioritize their risk mitigation efforts. Changing the mathematical approach to scores means that businesses can use their scorecards, and those of their vendors, more efficiently.

“Normalized” Universal Scoring: Historically, scoring was calculated by using a cohort benchmark, meaning companies were categorized by IP size and also by industry, and put into the relevant bucket, or “cohort.” Now, scores are calculated using a statistical baseline that includes all rated companies in the SecurityScorecard platform. This allows us to provide accuracy in scores by leveraging the unmatched number of companies we monitor. The new data-driven model utilizes a periodic “normalization” process to ensure all metrics used to calculate scores are statistically modeled across time to provide scoring accuracy. The benefit of this more advanced statistical model is that companies can now compare scores across any company, regardless of their industry or size.

More Robust Statistical Model: The Next-Gen scoring model accounts for variations in a company's size and issue volume by using logarithmic statistical analysis.

3 Month Calibration: A fundamental concept in Next-Gen scoring is the use of historical data to ensure effective calibration of the scoring model. Next-Gen leverages a trailing average of the past three months of data as the basis for normalization. Calibration is re-calculated monthly using the previous three months of data. A benefit of this approach is that it reduces the sensitivity of scoring from a major fluctuation of some aspect of the scoring data inputs (e.g., a large number of newly discovered assets in a company's digital footprint). This approach also allows SecurityScorecard to add new issues into scorecards as we proactively improve our signals. Another benefit of this approach in view of the evolving nature of cybersecurity threats is that it enables newly relevant security signals to be added and older issue types to be retired, as needed, so that security ratings always reflect the current threat landscape.

Removes "IP Buckets": As discussed previously, our traditional scoring model had a high reliance on industry and company size (i.e., digital footprint) as a key influence of score. Cohorts were determined based on alignment with similar companies of a similar size (i.e., "IP Buckets"). Now this model moves to a comparison based on a smooth distribution of company size vs. buckets. SecurityScorecard is confident this change translates to significant improvements across all three areas of accuracy, transparency, and fairness.

Issue Transparency in Factor Scores: NextGen scoring provides transparency of score calculations, enabling customers to have full visibility on the potential score impact from resolving issues.

NextGen Scoring Proves a Commitment to Transparent, Fair, and Accurate Security Ratings

SecurityScorecard's commitment to providing fair and clear ratings requires consistent improvement. Our new NextGen Scoring proves this commitment to enabling security through stable and user-friendly ratings in three ways.

Continued Commitment to Accuracy: Normalizing all companies delivers an unbiased security assessment and more advanced statistical techniques introduce additional confidence that our reporting is defendable and deterministic.

Enhanced Transparency: NextGen scoring enables new tools which allow organizations to better prioritize and address security concerns.

Improved Reporting Abilities: A key benefit of the normalized universal score is that it will greatly expand a user's ability to group and analyze companies against a broader range of business segmentation, including peer companies, partner companies, geographic regions or business unit boundaries.

How it Works

Many types of security issues scale with the size of the organization. Larger organizations typically have a larger "attack surface". More employees mean more devices to be protected and more servers mean more changes for an exposed port which should properly sit behind a firewall. Some issue types scale with the number of IPs. Other might scale with the number of related domains.

What is a Logarithm?

Business growth does not follow a straight line pattern nor a bell curve.
For example, a business rarely doubles its revenue every year. For example, a small business may be able to grow from $100,000 to $200,000 in profit during its second year then see growth to $325,000 profit in the third year.
Linear growth would be consistent profit growth per year, thus $100,000 in growth every year for the above example. Exponential growth would be consistently multiplying profit by the number of years, thus the third year should be $400,000 profits using 2 as a multiplier. Neither of these models represent realistic or sustainable business models. Logarithms, however, look at increases based on an exponent multiplier but also allowing mathematical accounts for the time it takes the company to grow.

Counting with Logarithms

Using Logarithms for Security Rating Scores

The best way to make meaningful measurements over such a large dynamic range is to use a logarithmic scale, where each increment corresponds to a multiple of 10. As noted above, the digital footprint of different organizations can vary from a single IP to hundreds of millions of IPs. This range spans more than eight orders of magnitude, or more than eight multiples of ten. Other common examples where a logarithmic scale is used to compare measurements spanning a wide dynamic range include the following:

- Richter scale for measuring earthquakes over 9+ orders of magnitude. For example, an earthquake of magnitude 8.0 on the Richter scale has 10x greater amplitude than one of magnitude 7.0.
- Decibel (or Bel) scale for measuring sound amplitude over 12 orders of magnitude.
- pH scale for measuring chemical acidity over 14 orders of magnitude.

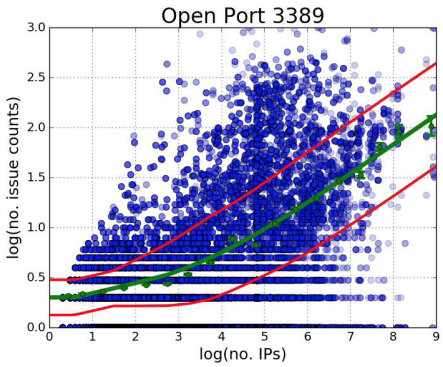
Signal Collection

SSC collects security signals across the IPv4 webspace on IPs that are controlled by millions of organizations. Accurate assignment of IPs and IP ranges to different organizations is controlled by our Attribution Engine, which combines publicly-available data about IP allocation with data collected by both broad and in-depth scanning to determine the digital footprint of an organization. . The Attribution Engine allows SSC to capture and identify the IPs associated with every entity scored on the SSC platform.

Signal Analysis

SecurityScorecard uses scatter plots to look for patterns in the collected signals information. Each dot represents a data point, meaning the more information gathered, the more accurate the patterns.

For each organization and each security issue – for example open port 3389 which corresponds to Microsoft's Remote Desktop Protocol – the number of occurrences of the issue type is captured. A scatter plot is generated in which every scored entity represents a point in a log-log plot of the logarithm of the number of issue counts vs. the logarithm of the number of IPs. A typical scatter plot will contains millions of data points, providing a large statistical "mass" for better accuracy and stability.



The scatter plot makes it easy to see where most data points exist as well as the location of outliers. A locally-weighted, nonparametric fitting algorithm is then applied to characterize the mean (green curve) and standard deviation (red curves) of the number of expected issue counts as functions of organization size. This calibration process is performed over a 3-month time interval to smooth out statistical fluctuations and is carried out for every issue type scored in the SSC platform.

The calibration process described above enables a reliable and stable statistical estimate to be calculated for a given organization and security issue, corresponding to how many standard deviations above or below the mean that organization is situated for the given security issue. In statistical parlance, this is known as a "z-score". For every issue type, SSC uses a "modified z-score", where z = 0 if no findings are present and z = 1 when the number of findings equals the mean or average for an entity with the same size digital footprint. The modified z-scores are calculated and updated daily for every entity monitored on the SSC platform.

This approach ensures inherently low score volatility meaning that the value is more stable. If an entity's digital footprint is unchanged and its issue counts are stable, then its security score will be rock solid.

Score Calculation

Issue types (i) are grouped into related categories or factors, such as Application Security or Network Security. For each entity (d) and factor (f), a raw factor score (RFS) is calculated as a weighted sum over the modified z-scores (zid), where the issue weights (wi) are severity based:

$$RFS_{fd} = \sum_{i \in f} w_i \times z_{id}$$

The raw factor scores are then scaled from 0 to 100, where 100 corresponds to a perfect factor score with no security findings and a score of 0 would indicate an abundance of security defects.
Finally, the total score is tabulated as the weighted average of the factor scores, where the individual factor weights are also based on severity. Since, in a security context, "a chain is only as strong as its weakest link", the total score also incorporates a non-linear weighting that amplifies the impact of a poor factor score. Consequently, one or more low factor scores will penalize the total score and drag it down by an additional amount. Letter grades based on a familiar grade-school rubric are assigned for easy interpretation in accordance with the table to the left.

Letter Grade	Numeric Score
A	90 - 100
B	80 - 89
C	70 - 79
D	60 - 69
F	0 - 59

Scoring on a Geometric Scale

A noteworthy consequence of using a logarithmic scale (rather than a linear scale) is that the dependence of score change on issue count change is geometric rather than arithmetic. This means, for example, that whenever issue counts are reduced by a fixed multiple, say a factor of two, the score will improve by a certain amount. (The exact amount will vary with issue type and entity size.) This is analogous to other logarithmic scales, such as with sound, where doubling the frequency of the pitch raises it by one octave.

Calibration and Cadence

SecurityScorecard's analysis found that individual industry information aligned with universal data. Looking at all organizations provides more information – and more information makes the patterns stronger. Thus, since the global information and the industry specific information are highly correlated, universal calibration allows for stronger estimates without losing the insight industry cohort reviews provided.

In NextGen scoring, a single universal or global calibration is carried out covering all entities in the platform rather than multiple calibrations for different industry subcategories. A universal calibration produces a more robust statistical estimate of means and standard deviations, and regression testing has demonstrated a high degree of correlation between universal calibration and calibration by industry ($R2 > 0.9995$).

To ensure statistical stability, the score calibration process described above is performed using daily measurements of security issues accumulated over a trailing 3-month time period and updated monthly. Averaging over a 3-month period mitigates potential error due to fluctuations in signal collection. Updating the calibrations on

a monthly cadence permits new issue types to be introduced and scored on a timely basis, which is important given the ever-evolving landscape of cybersecurity threats.

New Features and Additional Benefits

Score Impact: Scoring becomes more transparent as users can understand the quantitative impact on score for each issue type on their scorecard.

Score Planner: This new feature enables users to perform “what-if” scenarios to help prioritize and plan issue resolution. This feature leverages the increased transparency of how issues impact score, as discussed in the previous sections. There is also an option to use the SecurityScorecard optimization algorithm to design a customized remediation plan requiring the minimum number of issues to be resolved in order to achieve a user-specified target score.

Issue-level Event Log: Issue level Event Log introduces transparency around score changes by showing a clear record of what has happened in a scorecard and issues that impacted the score.

Summary

The foundation of NextGen scoring rests on sound statistical and analytical methods which provide new benefits and facilitate highimpact features. NextGen scoring and related features improve a user’s ability to answer important questions, like “Why did my score change?” or “How do I fix my score?”.