SecurityScorecard

Applying Machine Learning to Optimize the Correlation of SecurityScorecard Scores with Relative Likelihood of Breach

By Bob Sohval, PhD
Vice President and Fellow, Data Science

FACTORS

1. Application Security
2. Cubit Score
3. DNS Health
4. Endpoint Security
5. Hacker Chatter
6. Informational Leak
7. IP Reputation
8. Network Security
9. Patching Cadence
10. Social Engineering

Introduction

SecurityScorecard ratings provide a means for objectively monitoring the cybersecurity hygiene of organizations (including their vendors) and gauging whether their security posture is improving or deteriorating over time. The ratings are valuable for vendor risk management programs, determining risk premiums for cyber insurance, executive-level and board reporting, monitoring and resolving risks across your own attack surface (self-monitoring), and for assessing compliance with cybersecurity risk frameworks.

Cybersecurity ratings can be compared to financial credit ratings. Just as a poor credit rating is associated with a greater probability of default, a poor cybersecurity rating is associated with a higher probability of sustaining a data breach or other adverse cyber event(s).

SecurityScorecard provides security ratings on more than 12 million entities worldwide. The ratings score is a weighted average across more than 200 different types of measurements that span a cross-section of 10 cybersecurity factors, including Application Security, DNS Health, Endpoint Security and others, resulting in a composite score that is predictive of breach.

We recently conducted a large scale study, using Machine Learning (ML) to tune the weight of each measurement type, so that the total score is optimally correlated with the relative likelihood of incurring a data breach.

Materials and Methods

The analysis was carried out by backtesting over a 4 1/2 year period spanning 2019 through mid-2023. Reports of publicly disclosed breaches over this period were collected organically and from commercial and public data sources, including the Vocabulary for Event Recording and Incident Sharing (VERIS) Community Database, Vigilante, and US HHS, as well as inquiries to states attorneys general under the Freedom of Information Act. Breach reports that were attributed to employee theft, stolen laptops, or inadvertent disclosure were excluded from the analysis. From these sources, a cohort of 16,583 breaches was created.

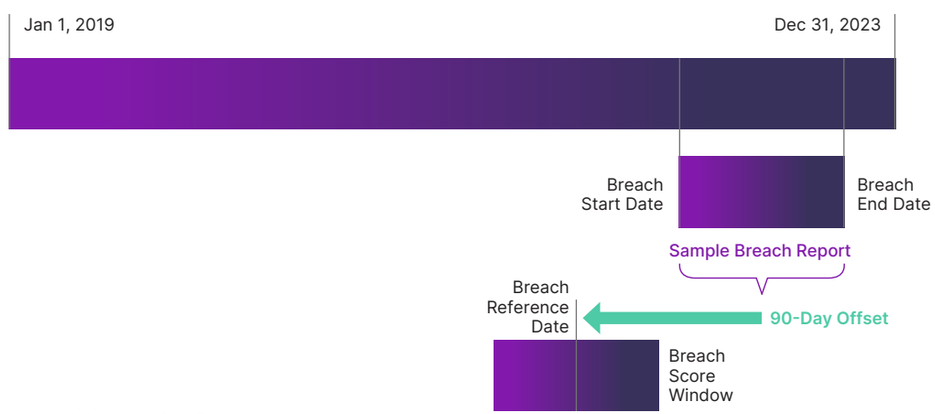
A second cohort of 16,583 non-breached organizations was created using stratified random sampling, drawing from a pool of followed vendors (excluding those who had sustained a breach) so that the distribution of digital footprint sizes for the non-breach cohort matched that of the breach cohort. This statistical procedure helped eliminate possible sources of systematic bias in the analysis.

The cyber threat landscape is dynamic. Over time, new cyber threats may emerge and SecurityScorecard's Collections Team may respond by introducing new measurement types to the platform. To accommodate this fact of life, the 4½ year time period in this study was divided into successive calendar quarters which could be independently evaluated. The stratified random sampling process described above was applied to each calendar quarter, ensuring that systematic bias was minimized.

In general, the exact date on which a breach occurred is not known with precision. Publicly disclosed breaches are typically reported by the affected organization with estimated begin- and end-dates. It is widely acknowledged that the elapsed time between the occurrence of a breach and its detection is typically several months. Based on data from the VERIS database, the median elapsed time from occurrence to discovery of a breach was found to be 90 days.

The effective date and measurement data for each breached organization were calculated as follows: the effective date is defined as the date 90 days prior to the halfway point between the reported begin- and end-dates. The 90-day offset accounts for the typical elapsed time between breach occurrence and detection. Since this is an estimated breach date, the breached organization's measurements were averaged over a window extending from 4 weeks prior to 4 weeks after the estimated breach date. If the breached organization was not scored during this window (for example, the organization was added to the platform and was first scored after the breach occurred), it was excluded from the analysis.

Timing Diagram for estimating breach date & score



Study Parameters

Evaluation Period 4.5 Years
Period Start January 1, 2019
Period End June 30, 2023
No. Data Breaches 16,583
No. Non-Breaches 16,583

1. http://veriscommunity.net/index.html

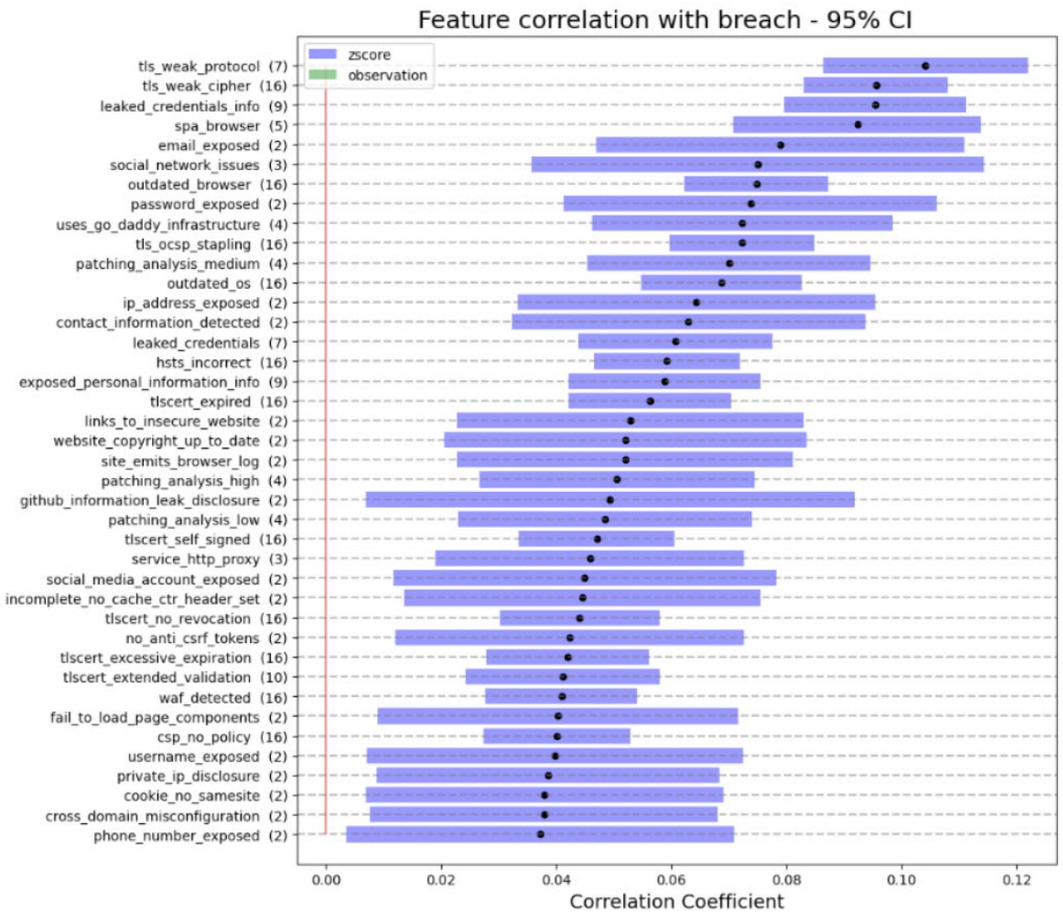
Identifying Measurements Associated with Breach

SecurityScorecard collects more than 200 different types of measurements to assess the cybersecurity posture of more than 12 million organizations worldwide. Different measurement types can reveal weaknesses or flaws in DNS configurations, the use of insecure protocols and outdated software, exposure to critical vulnerabilities, and others.

To determine which measurement types are most strongly associated with breach, the Data Science team systematically measured the correlation coefficient of each measurement type against breach, using the size-matched breach and non-breach cohorts. In addition, a chi square analysis was performed to potentially identify cybersecurity flaws that occur relatively rarely, but are statistically correlated with breach.

A portion of the results of the correlation study are shown in the chart below. The chart presents measurement types (features) in ranked order based on the magnitude of the correlation coefficient with breach. The vertical red line corresponds to zero correlation. The black dots correspond to the estimated value of the correlation coefficient, and the horizontal blue bars depict the 95% confidence intervals (CI) for each correlation coefficient measurement. The further the blue bar is situated to the right, the larger the magnitude of the correlation coefficient and the greater the statistical confidence that the given measurement type is truly correlated with breach.

The size of the 95% confidence intervals can vary in accordance with the quantity of the underlying data. The numbers in parentheses indicate the number of calendar quarters over which data for each measurement type had been collected at the time of the study. Typically, more recently added signals collected over fewer calendar quarters have larger “error bars”, reflecting less available data. However, the quantity of data has been properly accounted for in the statistical analysis.



Interpretation of Results

How do we understand these results?

Outdated browser and outdated OS appear in the upper portion of the ranked list of measurement types correlated with breach. These are both endpoint signals. While each of these signals can indicate a security risk by themselves (both are frequently updated with security patches), detection of these signals may indicate deeper concerns within the organization. Typically, a breach is characterized by a penetration event, followed by lateral movement within the infrastructure, and an escalation of privileges to access the treasured data. Arguably, the largest component in the attack surface of an organization is its employee base. If employees are not keeping their browsers uptodate, what other “bad behaviors” are they engaging in? Are they clicking on unsafe attachments? Are they falling prey to phishing emails? Have they even been trained in good cybersecurity practices? These endpoint signals may serve as proxies for inadequate cybersecurity hygiene.

The signals found to have the highest correlation with breach, perhaps surprisingly, relate to the strength of Transport Layer Security protocols, which are used to encrypt communications across a computer network to ensure secure transmission. Use of older and weaker ciphers and protocols may signal to potential hackers that the organization is not adhering to best practices and may be a ripe target for a cyber attack.

Poor Cyber Hygiene

The picture that begins to emerge from this analysis is analogous to the notion of a “safe driver”.

While driving above the speed limit or other moving violations may not directly cause an accident, insurance companies know that evidence of such unsafe behavior is statistically correlated with a higher incidence of automobile claims, and insurers typically charge higher premiums to offset the elevated risk associated with unsafe behavior.

Similarly, we observe a strong statistical correlation between “unsafe behaviors” — whether by the employee base or by security teams — and a higher incidence of data breach.

Correlation is Not Causation

In the current study over a large number of publicly disclosed breaches accrued over several years, it is not practical to determine the root cause of individual breaches. However, it is possible to make statistical observations across many breaches and the cyber profiles of breached and non-breached organizations, and to identify measurement types that occur statistically more prominently among breached organizations, and are not merely chance observations.

While correlation indeed is not the same as causation, this type of analysis can materially help risk managers identify organizations facing elevated statistical risk for an adverse cyber incident and take appropriate action to mitigate that risk.

Calculating the Score

SecruityScorecard reports both Total Score and Factor Scores.

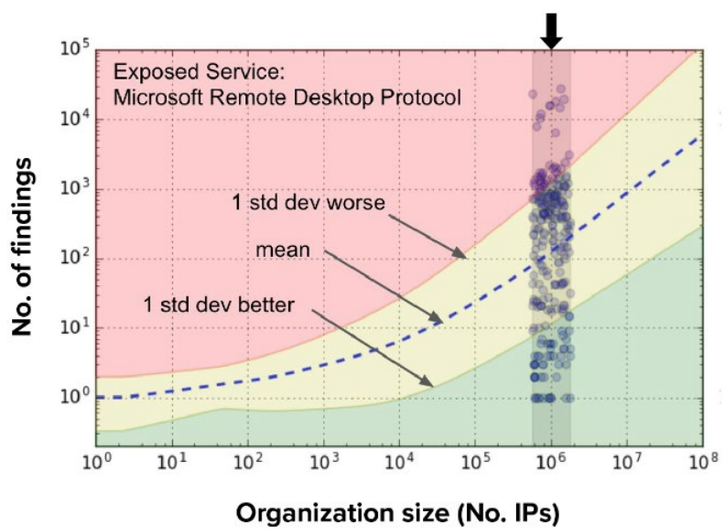
Factor Scores are calculated using a subset of measurement types that are relevant for the given factor. For example, Endpoint Security consists of measurements of Outdated Browsers and Outdated OSes.

While Factor Scores are calculated and presented on the scorecard, they are no longer used in the calculation of Total Score. Rather, Total Score is calculated directly from the observed measurements.

A challenge in the cybersecurity ratings space is how to level the playing field between organizations with different sizes of digital footprints. A "momandpop.com" might have only a handful of IP addresses, while a large organization with a sprawling infrastructure might be managing 100s of millions of IP addresses. Generally, larger infrastructures will be afflicted with a larger number and greater diversity of cybersecurity flaws compared to smaller ones.

SecurityScorecard addresses this challenge using a principled statistical approach, for which an example is shown below. For each measurement type (Remote Desktop Protocol in the example), the average number of findings for a given footprint size (dashed blue line in the figure) is determined by a non-parametric, data-driven approach using all 12+ million organizations scored on the platform. In the example shown here, each blue dot corresponds to a scored organization. Most have been removed for the sake of visual clarity. The yellow band depicts one standard deviation from the mean. Organizations in the red are at least one standard deviation worse than average and those in the green are at least one standard deviation better than average. The number of standard deviations from the mean is called a "z-score".

For each organization and each measurement type, SecurityScorecard calculates the z-score. Measurement level weights are determined based on their measured correlation coefficient with breach, as described above. Total Score and Factor Scores are calculated using a weighted sum of z-scores for the underlying measurements, and then rescaled to produce a final score in the range 0 to 100.



Validation

Total Scores for breached organizations, based on their estimated dates of breach, and Total Scores for non-breached organizations, drawn from the list of more than 500,000 companies followed by other organizations on the SecurityScorecard platform, were calculated as described on the previous page.

The relative breach likelihood ratio R(g) for grade g, where g = {A, B, C, D, F}, was calculated as follows:

$$R(g) = \frac{r(g)}{r(\mathrm{A})}$$

Where r(g) is the ratio of the number of breaches of organizations with grade g compared to the number of organizations (breach and non-breach) with grade g:

$$r(g) = \frac{n_{breaches}(g)}{n_{organizations}(g)}$$

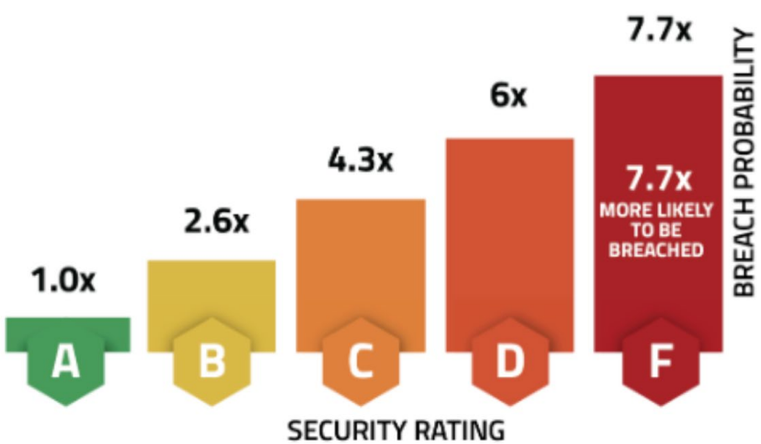
By definition, the relative breach likelihood ratio for a grade of A is 1.0. If poor grades are correlated with greater breach likelihood, then the breach likelihood ratio R(g) should be greater than 1.0 for worse grades.

Results

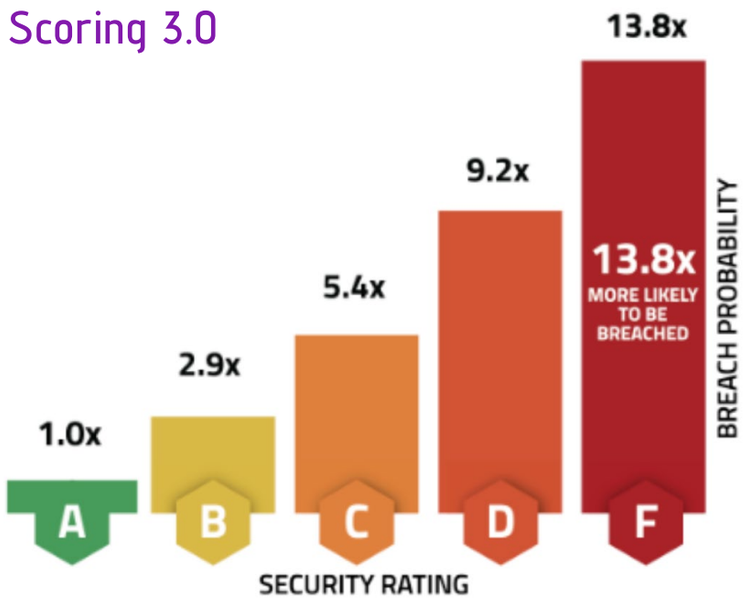
As depicted in the figure below, relative breach likelihood was found to increase monotonically as the score and associated grade worsens, culminating with the statistical result that an organization with a grade of 'F" is 13.8x more likely to sustain a breach than an organization with an 'A grade.

This represents a 79% improvement in the breach likelihood ratio compared to the previous scoring methodology.

Scoring 2.0



Scoring 3.0



Conclusion

Companies managing the cyber risk of a portfolio of organizations — for example as part of a vendor risk management program — may use these results to make more informed risk assessments. While actual risk values will likely vary depending on the precise composition of a given portfolio, the results from the present analysis are believed to be representative, and can assist cybersecurity practitioners and risk managers to more accurately assess breach risk.

The application of machine learning to further improve the correlation of SecurityScorecard grades with breach likelihood by nearly 80%, backtested and validated over a 4.5-year period, constitutes a significant milestone and important advance in the maturity and accuracy of cybersecurity ratings.

About SecurityScorecard

Funded by world-class investors including Evolution Equity Partners, Silver Lake Waterman, Sequoia Capital, GV, Riverwood Capital, and others, SecurityScorecard is the global leader in cybersecurity ratings with more than 12 million companies continuously rated.

Founded in 2013 by security and risk experts Dr. Aleksandr Yampolskiy and Sam Kassoumeh, SecurityScorecard's patented rating technology is used by over 30,000 organizations for enterprise risk management, third-party risk management, board reporting, due diligence, cyber insurance underwriting, and regulatory oversight. SecurityScorecard is the first cybersecurity ratings company to offer digital forensics and incident response services, providing a 360-degree approach to security prevention and response for its worldwide customer and partner base.

SecurityScorecard continues to make the world a safer place by transforming the way companies understand, improve and communicate cybersecurity risk to their boards, employees and vendors. Every organization has the universal right to their trusted and transparent **Instant SecurityScorecard** rating. For more information, visit **securityscorecard.com** or connect with us on **LinkedIn**.

GET YOUR SCORE

Want to receive an email with your company's current score, please visit instant.securityscorecard.com.

Get Started

SecurityScorecard

SecurityScorecard.com
info@securityscorecard.com