



# Cyber Defender Strategies: What Your Vulnerability Assessment Practices Reveal

# CONTENTS

<b>Executive Summary</b>	3
Findings	3
Recommendations	4
<b>Introduction</b>	4
Vulnerability Assessment Objectives	5
<b>Vulnerability Assessment Key Performance Indicators and Maturity</b>	7
<b>Analysis</b>	10
Vulnerability Assessment KPIs by Style	11
General VA Style Distribution	12
VA Style Distribution by Geography	13
Key Findings	13
VA Style Distribution by Employee Count	14
VA Style Distribution by Licensed Asset Count	14
VA Styles by Industry	15
<b>Conclusion</b>	16
Findings Summary	17
<b>Recommendations for VA Maturity Levels</b>	18
<b>Appendix</b>	19
Methodology	19
Archetypal Analysis	19
References	20
Acronyms	20

# I. EXECUTIVE SUMMARY

In this report we analyze real-world end-user vulnerability assessment (VA) behavior using a machine learning (ML) algorithm to identify four distinct strategies, or “styles.” These are based on five VA key performance indicators (KPIs) which correlate to VA maturity characteristics.

This study specifically focuses on key performance indicators associated with the Discover and Assess stages of the five-phase Cyber Exposure Lifecycle. During the first phase – Discover – assets are identified and mapped for visibility across any computing environment. The second phase – Assess – involves understanding the state of all assets, including vulnerabilities, misconfigurations, and other health indicators. While these are only two phases of a longer process, together they decisively determine the scope and pace of subsequent phases, such as prioritization and remediation.

The actual behavior of each individual enterprise in the data set, in reality, exhibits a mixture of all VA Styles. For the purposes of this work, enterprises are assigned to the specific style group with which they most closely align. We provide the global distribution of VA Styles, as well as a distribution across major industry verticals.

## FINDINGS

- Enterprises conducting VA fall into four distinct VA Styles, ordered by maturity: Diligent, Investigative, Surveying and Minimalist.
  - The Diligent style represents the highest maturity, yet constitutes only five percent of all enterprises in the data set.
  - The Investigative style represents a medium to high maturity, with 43 percent of enterprises following this style.
  - The Surveying style, with a representation of 19 percent in the data set, corresponds to a low to medium maturity.
  - The Minimalist style represents the lowest maturity and constitutes 33 percent of all enterprises in the data set.
- The hospitality, transportation, telecommunications, electronics and banking industries had the highest proportion of the mature Diligent style.
- The utilities, healthcare, education and entertainment industries had the highest proportion of the low-maturity Minimalist style.
- The utilities industry had the highest proportion of the low-maturity Minimalist style overall.
- The distribution of VA styles by geographical region shows no noteworthy variation.

## II. INTRODUCTION

The cybersecurity community is heavily focused on what attackers are doing. While threat intelligence and vulnerability research is invaluable, it only represents one side of the equation. Far less research has been dedicated to how defenders are responding.

There is a wealth of qualitative data available on what end users are doing, primarily derived from surveys. The reliability of survey data is dependent on the knowledge and honesty of participants. Results can be skewed by cognitive biases and lack of awareness. What someone believes they are doing is not always the same as what they are actually doing, especially when practical realities come into play. Quantitative research based on end-user behavior and telemetry data provides a more reliable basis for determining the true state of general VA maturity.

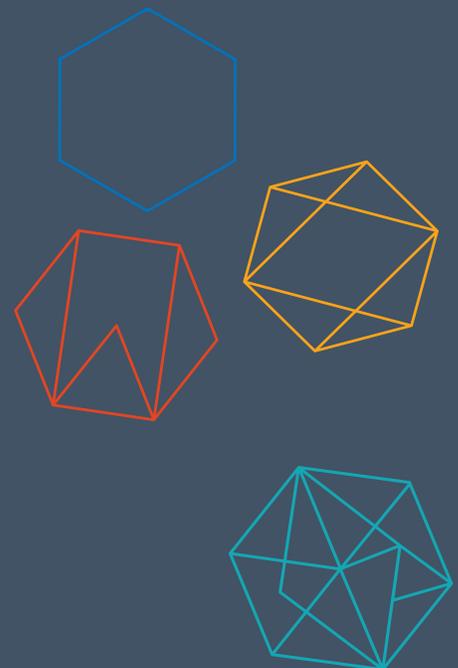
In our last report, [“Quantifying the Attacker’s First-Mover Advantage,”](#) we discovered attackers generally have a median seven-day window of opportunity during which they have a functional exploit available to them, before defenders have even determined they are vulnerable. The resulting seven-day gap is directly related to how enterprises are conducting VA.

In this study, we analyze real-world VA telemetry data to group end users into segments and identify four distinct strategies, or “styles,” of VA. Further analysis focuses on the distribution of these four VA Styles across industries.

To classify the VA Styles, we applied a machine learning algorithm called archetypal analysis (AA) to real-world scan telemetry data from more than 2,100 individual organizations in 66 countries and just over 300,000 scans during a three-month period from March to May 2018. AA identifies a number of idealized/archetypal VA behaviors within this data set. Organizations are assigned to groups defined by the archetype they are most similar to. This does not mean each organization in a group behaves exactly like the archetype. Rather, it means that, of the four archetypes, they are most similar to the archetype which defines that grouping. The scanning behavior styles described in this report are based on these four archetypes.

## RECOMMENDATIONS

- Evaluate your own vulnerability assessment maturity based on our five critical VA KPIs: Scan Frequency, Scan Intensity, Authentication Coverage, Asset Coverage and Vulnerability Coverage.
- Identify your current VA Style and compare yourself to industry peers.
- Follow the recommendations for your style to determine the KPIs you need to improve to move your maturity to the next level.



## VULNERABILITY ASSESSMENT OBJECTIVES

This study specifically focuses on key performance indicators (KPIs) associated with the Discover and Assess stages of the five-phase Cyber Exposure Lifecycle. During the first phase – Discover – assets are identified and mapped for visibility across any computing environment. The second phase – Assess – involves understanding the state of all assets, including vulnerabilities, misconfigurations and other health indicators. While these are only two phases of a longer process, together they decisively determine the scope and pace of subsequent phases, such as prioritization and remediation.

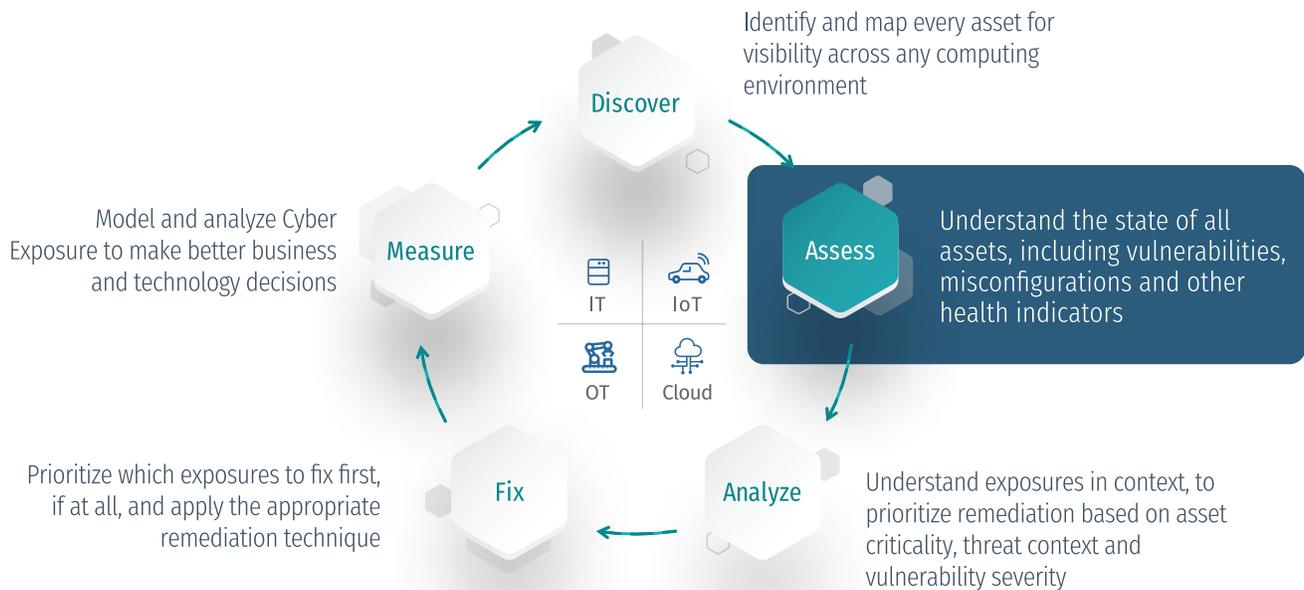


Figure 1: Tenable's Cyber Exposure Lifecycle

Vulnerability Assessment has traditionally been conducted by deploying a “scanner” to assess assets remotely over the network, interrogating any open ports and available services to see if they are vulnerable.

To accommodate diverse and complex use cases, and to cover emerging technologies, Vulnerability Assessment has expanded beyond pure dynamic remote scanning. Modern VA supports conducting assessments using local agents, by passive network monitoring, and by integrating with diverse third-party technologies – such as enterprise mobility management suites (EMM), hypervisors and Infrastructure-as-a-Service (IaaS) platforms – to gather additional data about vulnerability and asset state.

Authenticated scanning, where credentials are used to gain a more thorough and reliable view of an asset, has also become a staple in the vulnerability manager's toolbox. Additionally, modern VA solutions support the centralized management of a tiered and heterogeneous scanning architecture, permitting the scheduling of scans, distribution of larger assessments across a pool of scanners, and the creation and customization of use-case specific scan configuration profiles for individual asset groups, business units or threat scenarios.

Together, these capabilities provide the technological foundation for VA, but it is how they are used that ultimately decides the effectiveness of VA. The general objectives of an effective Vulnerability Assessment process are summarized below:

- Scan sufficiently to fulfill regulatory requirements.
- Scan as frequently as possible to minimize the length of time in which a critical vulnerability may reside in your environment without your knowledge, and to obtain up-to-date benchmarking and risk scoring intelligence.
- Gain as much visibility of critical vulnerabilities on assets as possible, beginning with unauthenticated remote assessments, and increasingly progressing to using authentication or a local agent to gain a system-side view as well.
- Assess as much of the infrastructure as possible, extending across all deployed assets, technologies and applications, to reduce the available attack surface an adversary can target.
- Leverage customized scan templates to tailor assessments to specific asset groups, business units and use cases, to reduce scan overheads and false positives and to limit unnecessary complexity.

In practice, many enterprises weigh each of these objectives differently and fulfill them to varying degrees. Technological debt, resource availability, risk appetite and business requirements are all major factors influencing VA maturity.

Measuring VA maturity is more art than science. There are many competing Information Security Management frameworks and compliance regimes, each with its own views on maturity. Below for example is how Gartner defines Vulnerability Assessment maturity in its [Vulnerability Management Maturity Model](#)<sup>1</sup>. Further on in this report, we will illustrate how the VA Styles align to Gartner’s model.

LEVEL	VA	REMEDIATION	MITIGATION	METRICS AND REPORTS
1	No repeatable VA; rare ad hoc VA by a consultant	Occasional patching of OS; default automatic patching (if any); no application patching; no overall remediation and mitigation planning	No mitigation	None
2	Compliance-driven unauthenticated scanning for external systems	Compliance-mandated remediation cycle; minimum automation	Ad hoc mitigation	Compliance reporting
3	Compliance-driven unauthenticated scanning	Compliance-mandated and some risk-based remediation	Network mitigation via NIPSSs and firewalls	Compliance reporting with some remediation progress reporting

<sup>1</sup>Gartner, A Guidance Framework for Developing and Implementing Vulnerability Management, Augusto Barros, Anton Chuvakin, 22 June 2017

LEVEL	VA	REMEDIATION	MITIGATION	METRICS AND REPORTS
4	A mix of authenticated and unauthenticated VA scanning; select systems' Secure Configuration Assessment (SCA)	VA and remediation logically connected; consensus remediation planning for risk reduction; mature process for validation of fixes	Network and endpoint mitigation; careful mitigation tracking	Compliance reporting, progress reports and risk-based reports; hotspot analysis
5	Comprehensive VA and SCA; authenticated scanning and near universal system coverage, including emerging IT environments	Tight integration of remediation, mitigation and monitoring; automated remediation and risk-based prioritization; analytics-driven decision making for remediation; automated validation of remediation actions	Risk-driven mitigation that is linked to remediation and security monitoring	Risk-based reporting, trending and metrics; continuous improvement based on the measures

Figure 2: Gartner's Vulnerability Assessment Maturity Levels

### III. VULNERABILITY ASSESSMENT KEY PERFORMANCE INDICATORS AND MATURITY

Our data model analyzes distinct vulnerability assessment performance indicators derived from VA behavioral telemetry data. These KPIs correspond to VA maturity. The table below details the KPIs we chose to measure to determine maturity:

SCAN KPI	WHAT IT MEASURES
Scan Frequency	<p>Scan Frequency measures how often an enterprise conducts assessments, based on the average length of time between days when a scan ran (scan day). A higher frequency means fewer days between assessments, and consequently means critical vulnerabilities can be identified faster.</p> <p><b>Low = Scans every week, every month, or even less often</b>  <b>Moderate = Scans every three to seven days</b>  <b>High = Scans more frequently than every three days</b></p>
Scan Intensity	<p>Scan Intensity measures how many different scans are launched on a given scan day. A higher Scan Intensity indicates an organization is executing multiple scans, whether to distribute a large scan across multiple scanners, or because they are using differentiated and customized scan templates to cover different asset groups, technology families, or use cases.</p> <p><b>Low = One scan on a given scan day</b>  <b>Moderate = Between one and six scans on a given scan day</b>  <b>High = More than six scans on a given scan day</b></p>

SCAN KPI	WHAT IT MEASURES
<b>Authentication Coverage</b>	<p>Authentication Coverage (whether using credentials or local agents) is a measure of the assessment depth. Unauthenticated assessments only provide a very limited and partial view, and yield more false negatives than credentialed scanning.</p> <p><b>Low = Less than 30 percent of scans include authentication credentials</b>  <b>Moderate = 30 percent to 70 percent of scans include authentication credentials</b>  <b>High = More than 70 percent of scans include authentication credentials</b></p>
<b>Asset Coverage</b>	<p>Asset Coverage measures the proportion of the licensed assets scanned in a 90-day period. This is an important metric, as a low asset coverage may not be intended, but rather a consequence of misconfiguration or network routing issues.</p> <p><b>Low = Less than 30 percent of all licensed assets are assessed over a 90-day period</b>  <b>Moderate = 30 percent to 70 percent of assets are assessed over a 90-day period</b>  <b>High = More than 70 percent of assets are assessed over a 90-day period</b></p>
<b>Vulnerability Coverage</b>	<p>Vulnerability Coverage measures the proportion of total vulnerability plugins used in a 90-day period. This indicates the overall comprehensiveness of assessments in covering diverse technologies and vulnerability families. While it seems counterintuitive, a very high vulnerability coverage does not necessarily indicate a higher level of maturity. There are a variety of vulnerability detection plugins covering everything from mainstream to exotic technologies, so an excessively high vulnerability coverage in conjunction with only a single recurring scan indicates assessment is being conducted indiscriminately and without any customization. A high maturity approach will utilize a broad mix of vulnerability plugins to be able to cover all of the technologies an enterprise may have deployed. These technologies will be selected based on existing and specific asset demographics, and used in targeted scan profiles. Gratuitous vulnerability plugin selection adds overheads which reduce efficiency and affect scan duration, and can potentially increase the rate of false positives while introducing unnecessary complexity.</p> <p><b>Targeted = Less than 25 percent of all available vulnerability plugins</b>  <b>Comprehensive = 25 percent to 75 percent of all available vulnerability plugins</b>  <b>Untargeted = More than 75 percent of all available vulnerability plugins</b></p>

Figure 3: Scan Behavior KPIs used in the analysis

For reference, we approximate our VA Maturity KPIs to Gartner’s VA Maturity Model in the table below.

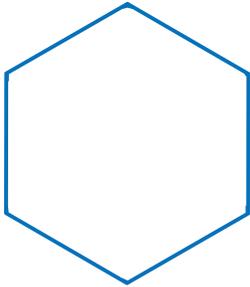
LEVEL*	CHARACTERISTICS	SCAN FREQUENCY	SCAN INTENSITY (PER DAY)	AUTHENTICATED SCANNING	ASSET COVERAGE	PLUGIN COVERAGE
1	*	*	*	*	*	*
2	Compliance-driven unauthenticated scanning for external systems	Low	Low	None	Low	Untargeted
3	Compliance-driven unauthenticated scanning	Moderate	Moderate	Low	Moderate	Untargeted
4	A mix of authenticated and unauthenticated VA scanning; select systems’ SCA	Moderate	Moderate	Moderate	Moderate	Targeted
5	Comprehensive VA and SCA; authenticated scanning and near universal system coverage, including emerging IT environments	High	High	High	High	Comprehensive

Figure 4: VA Maturity KPIs and Gartner’s VM Maturity Model

\*Level 1 indicates no repeatable VA is being conducted, and is therefore not included in the above table.

## IV. ANALYSIS

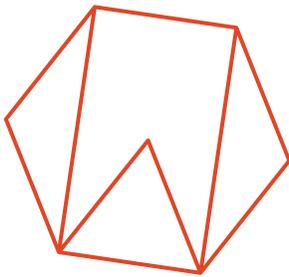
Our analysis resulted in four distinct Vulnerability Assessment Styles, or strategies, described below:



### THE “MINIMALIST” STYLE LOW MATURITY

The Minimalist executes bare minimum vulnerability assessments as required by compliance mandates.

- Scans every week, every month or even less often
- Executes a single scan at a time
- Authenticates little
- Partial asset coverage
- Leverages a single, comprehensive scan template



### THE “SURVEYING” STYLE LOW TO MEDIUM MATURITY

The Surveyor conducts frequent broad-scope vulnerability assessments, but focuses primarily on remote vulnerabilities.

- Scans every three days or less
- Executes a single scan at a time
- Authenticates little
- High asset coverage
- Leverages a single, comprehensive scan template



### THE “INVESTIGATIVE” STYLE MEDIUM TO HIGH MATURITY

The Investigator executes vulnerability assessments with a high maturity, but only assesses selective assets.

- Scans weekly or less
- Executes distributed or use-case specific scans
- Authenticates every scan
- Partial asset coverage
- Leverages a variety of streamlined, targeted scan templates



### THE “DILIGENT” STYLE HIGH MATURITY

The Diligent conducts comprehensive vulnerability assessments, tailoring scans as required by use case, but only authenticates selectively.

- Scans every three days or less
- Executes many segmented or differentiated scans
- Authenticates selectively
- High asset coverage
- Leverages distinct scan templates for different use cases

Figure 5: The Four VA Styles

The radar chart below shows where the four VA scanning behavior styles fall on the maturity scale for each of our five KPIs. The Minimalist style immediately sticks out, showing a low maturity level across all KPIs. The Diligent style is also noticeable, showing a high maturity across four out of five KPIs. The Investigative style shows a peak for Authentication Coverage, deviating from the moderate maturity displayed for the remaining KPIs. The Surveying style draws a trapezoid, displaying an uncharacteristic mix of low and high maturity in the KPIs.

## VULNERABILITY ASSESSMENT KPIs BY STYLE

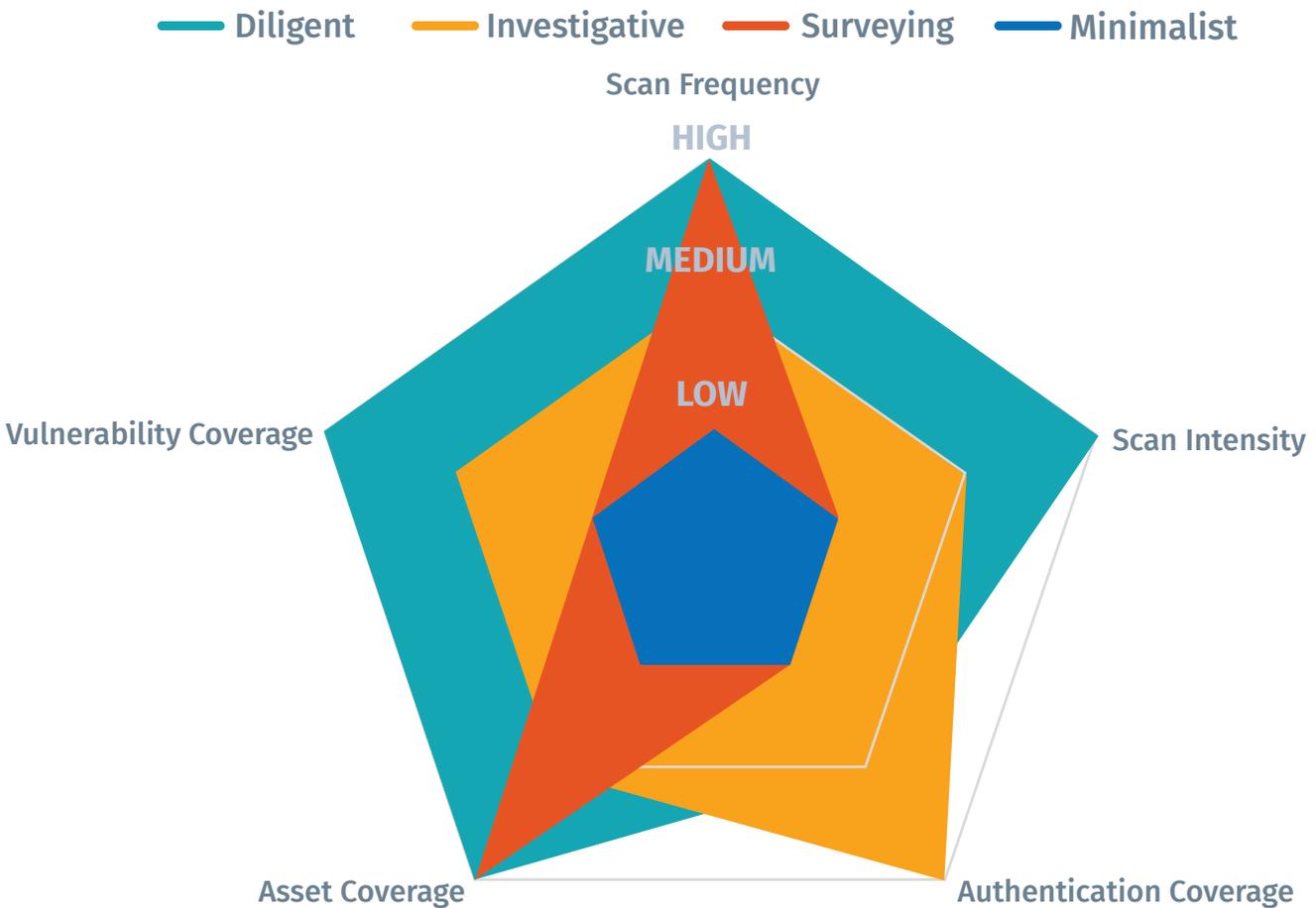


Figure 6: VA KPIs by Style

Our analysis indicates the reality of VA maturity is more nuanced than imagined by traditional frameworks. The heatmap in Figure 7 shows maturity doesn't improve linearly across the five KPIs measured.



Figure 7: VA KPIs by Style Heatmap

### GENERAL VA STYLE DISTRIBUTION

The chart below shows the general distribution of VA scanning styles across all enterprises included in the data set:

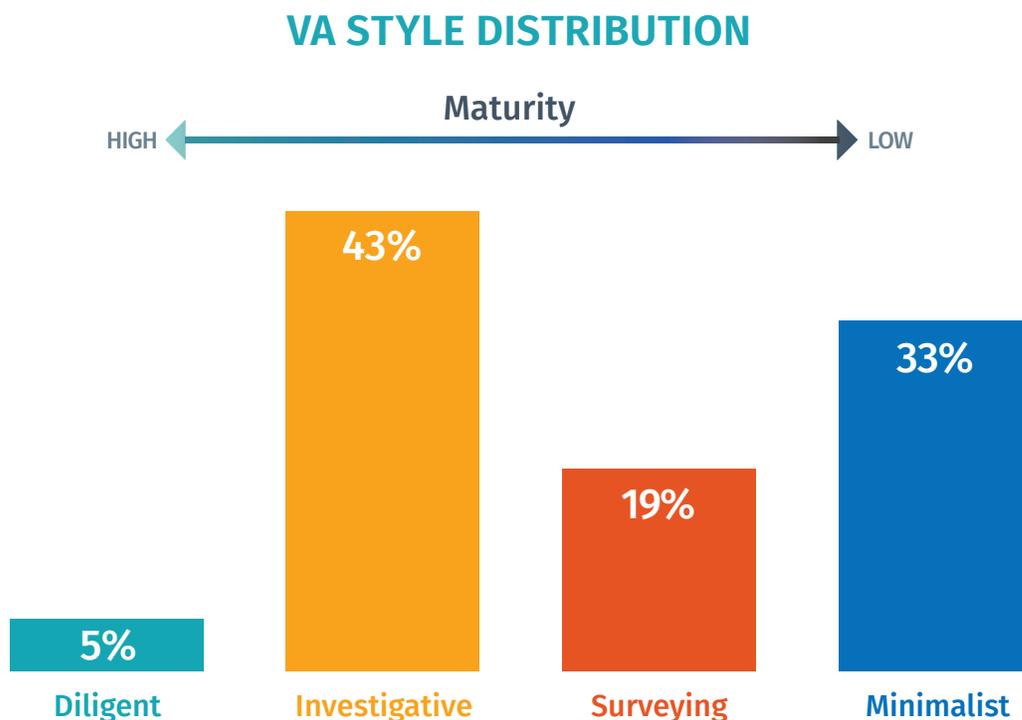


Figure 8: Overall VA Style Distribution

## VA STYLE DISTRIBUTION BY GEOGRAPHY

When we broke down the style distribution based on geographic regions, we were surprised to see very little variation between the three regions. Our conclusion is that, due to shared supply chains, the effects of globalization and the associated international trade norms, standards and regulations – as well as the relatively universal objectives of vulnerability management – geographical variations are less pronounced than anecdotal evidence suggests. We are planning future research on whether the differences are more pronounced on a national basis.

### Region

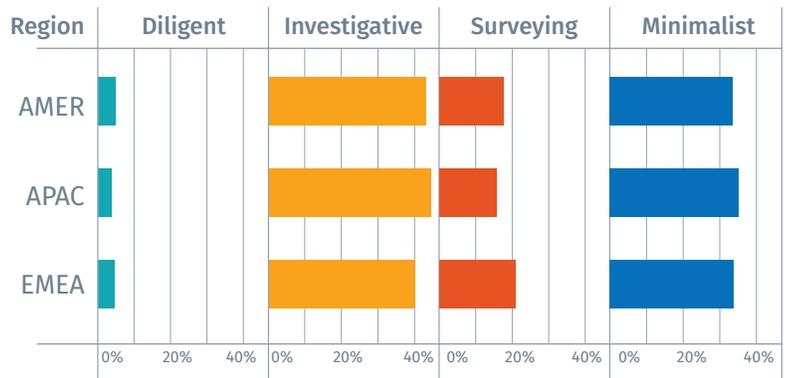


Figure 9: Style Distribution by Geography

## KEY FINDINGS



Only **5%** of enterprises follow the **Diligent style** and are at a higher level of maturity, displaying a high assessment frequency, comprehensive asset coverage, and targeted, customized assessments.

**43%** follow the **Investigative style**, indicating a medium to high maturity. These display a good scan cadence, leverage targeted scan templates, and authenticate most of their assets.

**19%** of enterprises follow the **Surveying style**, placing them at a low to medium maturity. Surveyors conduct broad scope assessments, but with little authentication and little customization of scan templates.

**33%** of enterprises are at a low maturity, following the **Minimalist style** and conducting only limited assessments of selected assets.

## VA STYLE DISTRIBUTION BY EMPLOYEE COUNT

Breaking down the style distribution by organization size based on number of employees shows a progressive increase in the more mature Diligent style as enterprises get larger.

The common wisdom is that cybersecurity maturity increases as an organization grows, and the data bears this out, but this does not seem to be a tide that lifts all boats. The percentage of Minimalist style followers, the least mature, is 30 percent to 40 percent in large enterprises with 5,000 employees and more. Also of note, the proportion of organizations engaged in the second least-mature style, Surveying, stays relatively constant across organization sizes.

Employee Range

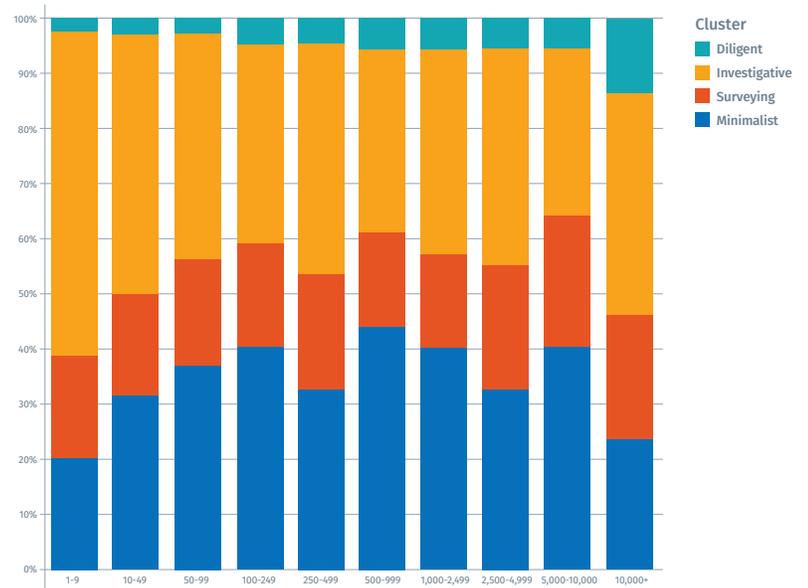


Figure 10: VA Style Distribution by Employee Count

## VA STYLE DISTRIBUTION BY LICENSED ASSET COUNT

We also drilled down into the style distribution based on licensed asset count. The biggest surprise for us was the correlation between licensed asset count and a higher proportion of the most mature Diligent style. Our expectation was an increase in maturity up to a certain count of licensed assets, and then a decrease due to increasing complexity of managing assets at scale and volume. Another interesting data point was that the least-mature Minimalist style peaked at a licensed asset count of between 200 and 499 assets.

Licensed Assets

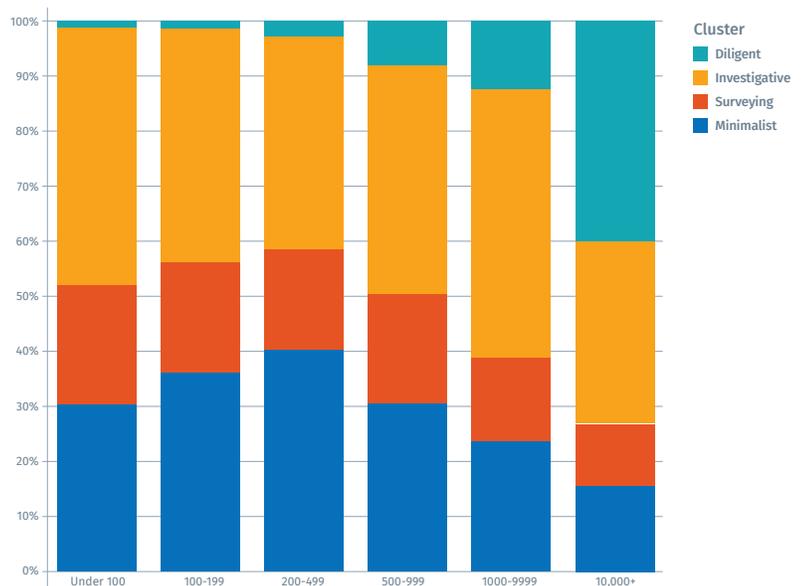


Figure 11: VA Styles by Asset Count

## VA STYLES BY INDUSTRY

While a breakdown by geography yielded little variation, breaking the styles down by industry vertical shows a much wider spread.



Figure 12: VA Styles by Industry

The industry to which an enterprise belongs appears to have a much greater effect than asset count on VA maturity and behavior, with some industries displaying a dominant style. Here's what we discovered:

	<b>The hospitality, transportation, telecommunications, electronics and banking industries had the highest proportion of the mature Diligent style.</b>
	<b>The utilities, healthcare, education and entertainment industries had the highest proportion of the low-maturity Minimalist style. The utilities industry had the highest proportion of the low-maturity Minimalist style overall.</b>
	<b>The medium to high maturity Investigative style is noticeably dominant in the engineering industry.</b>
	<b>The engineering and utilities industries show no representatives who follow the mature Diligent style.</b>

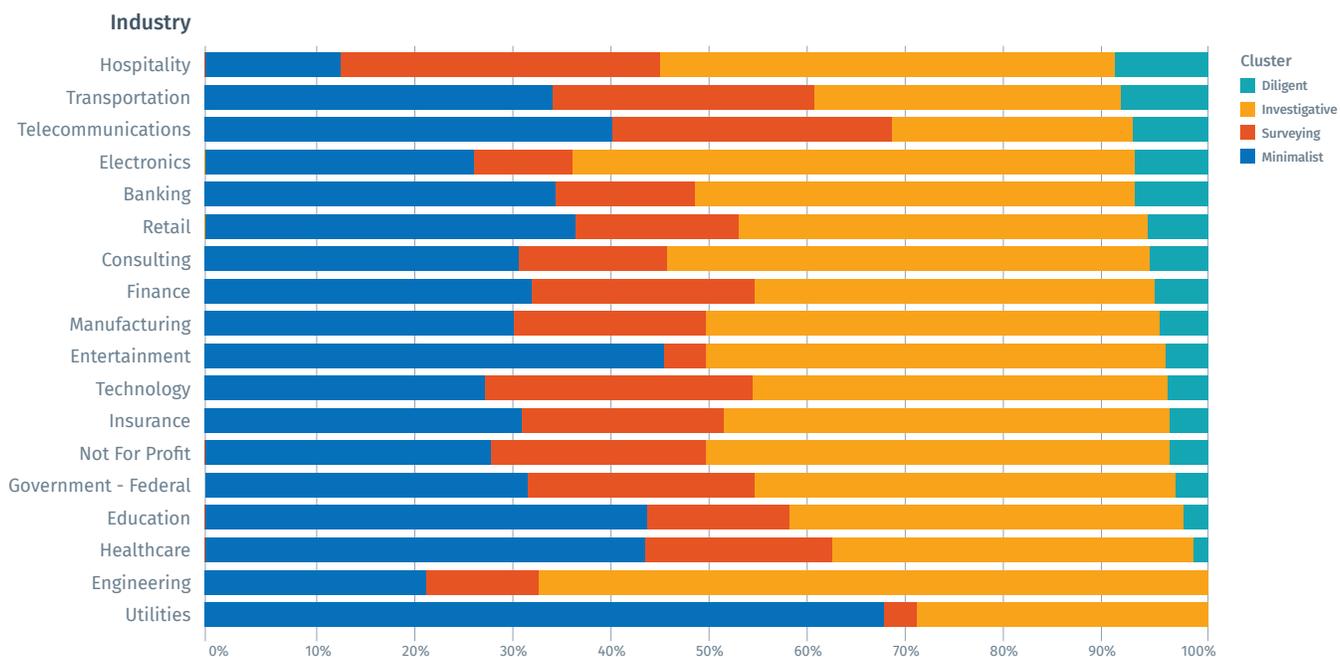


Figure 13: VA Styles by Industry and Maturity

When we sort the industries by the most mature Diligent style in the chart above, we see that, surprisingly, the hospitality industry has the largest proportion of Diligent style followers.

## V. CONCLUSION

Managing vulnerabilities and Cyber Exposure at scale under real-world conditions can feel like trying to repair a running engine in a car while driving down the highway at 70 miles per hour.

Despite this, five percent of enterprises are following the mature Diligent style, improving as company size increases. Diligent enterprises are acting strategically, scanning tactically and include most of their asset population in the scope of their vulnerability assessment program.

It is also promising to see 43 percent of enterprises in the data set are following the Investigative style, displaying a mix of mainly medium and some high maturity across the KPIs we measured. When we consider the challenges involved in managing vulnerabilities, getting buy-in from management, cooperating with disparate business units such as IT operations, maintaining staff and skills, and the complexities of scale, this is a great achievement and provides a solid foundation upon which to mature further.

On the other hand, 19 percent of the enterprises in the data set are most closely aligned with the Surveying style, with primarily low maturity across three of the KPIs, and high maturity across two. When we carefully consider the specific KPIs with a high maturity, we see Surveyors running regular broad scope assessments, but with little depth. This style will give a simpler baseline of what a remote attacker would see, but Surveyors must begin tailoring assessments for specific asset types and, most importantly, expand authentication coverage to gain a holistic view of their security posture.

Lastly, we see that 33 percent of the enterprises in the data set are following the low-maturity Minimalist style. That represents a lot of enterprises which are exposed to risk and still have some work to do, with critical decisions to make on which KPIs to improve first. Fortunately, the foundation for maturing their vulnerability management program is already in place.

## FINDINGS SUMMARY

- Only 5 percent of enterprises display high maturity characteristics.
- Compare this to the 33 percent following a low-maturity style.
- The style with the highest proportion of followers (43 percent) is the Investigative style, displaying a moderate maturity with high-maturity elements.
- Nearly half of all enterprises display mature characteristics.
- Conversely, nearly half are conducting VA at a medium- to low-maturity level.
- Surprisingly, there was very little geographical variation on the distribution of styles. Common wisdom states that differences in geographical business practices and regulations impact how companies conduct security. We plan to follow up with future research on why the differences do not seem as pronounced as expected.
- Company size had a greater impact on the distribution of styles and associated maturity.
- The distributions become more pronounced when we drill down into specific verticals:
  - The utilities, healthcare, education and entertainment industries had the highest proportion of the low-maturity **Minimalist style**.
  - The hospitality, transportation, telecommunications, electronics and banking industries had the highest proportion of the mature **Diligent style**.
  - The utilities industry had the highest proportion of the low-maturity **Minimalist style** overall.
  - Engineering, electronics and entertainment all had a noticeable bias for the **Investigative style**.

## RECOMMENDATIONS FOR VA MATURITY LEVELS

We provide these high-level recommendations for each style to help your organization improve VA maturity.



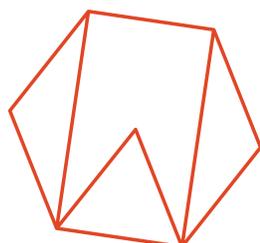
### DILIGENT STYLE

- Expand authenticated scanning (credential or agent-based) beyond select assets and technologies.
- Begin including non-traditional technologies in the scope of your Vulnerability Management program, such as web, cloud, virtual and mobile assets.



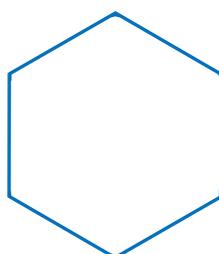
### INVESTIGATIVE STYLE

- Extend asset coverage to the broader organization, not just to select assets.
- Increase the scan frequency to minimize the time it takes to become aware of and respond to critical vulnerabilities.
- Expand the usage of customized scan templates focusing on specific technology families and for specific use cases, for example for exploitable vulnerabilities.



### SURVEYING STYLE

- Expand the use of credentials and agents for authenticated scanning to achieve a deeper and more reliable view of an asset's vulnerabilities.
- Leverage customized scan templates focusing on specific technology families and for specific use cases, such as exploitable vulnerabilities.
- Begin leveraging distributed scanning to load balance assessments across multiple scanners and reduce scan duration.



### MINIMALIST STYLE

- Reduce the number of days between regular assessments.
- Extend asset coverage to exposed and critical asset groups and business units.
- Leverage credentials or agents for authenticated scanning to gain a deeper and more reliable view of an asset's vulnerabilities.
- Begin leveraging distributed scanning to load balance assessments across multiple scanners and reduce scan duration.

## VI. APPENDIX

### METHODOLOGY

This research study analyzed vulnerability assessment telemetry data from more than 2,100 customers and 300,000 scans over a three-month period from March to May 2018. The data was classified using Archetypal Analysis, a machine learning algorithm for finding a small number of pure types or archetypes in a data set.

We use anonymized telemetry data collected from our Tenable.io platform in accordance with our end-user license agreement (EULA) to research trends and topics fundamental to cyber security. We do not use telemetry data from other Tenable products, like Nessus or SecurityCenter in our research and related reports.

The analysis focused on five key telemetry metrics:

SCAN BEHAVIOR CHARACTERISTIC	DESCRIPTION
<b>Scan Frequency</b>	A scan day is a day on which at least one scan was conducted. Average interval between scan days captured scan frequency.
<b>Scan Intensity</b>	The average number of different scans on a given scan day.
<b>Authentication Coverage</b>	The percentage of scans where credentials were provided and at least one asset was successfully authenticated against.
<b>Asset Coverage</b>	The proportion of assets scanned in 90 days compared to the total number of licensed assets.
<b>Vulnerability Coverage</b>	The proportion of the total number of available vulnerability plugins used in a 90-day period.

### ARCHETYPAL ANALYSIS

Archetypal Analysis is a method for finding a number of pure types or archetypes in a data set. The algorithm was introduced to the machine learning literature by Cutler and Breiman (1994). The goal of AA is to identify a number of archetypes that capture some idealized behaviors within the data set. The number of archetypes should be much less than the number of observations in the data set.

The archetypes are identified such that each observation can be well represented by some weighted combination of these archetypes. The archetypes themselves are constrained to be weighted combinations of the observations. In this way the archetypes are guaranteed to make physical sense. This is not the case with other techniques, such as principal component analysis (PCA), where impossibilities such as negative lengths can arise.

Since the archetypes are weighted combinations of the observations and the observations are weighted combinations of the archetypes, model fitting boils down to estimating two sets of weights. This is usually achieved via an iterative least squares algorithm. For a thorough exposition on this model fitting algorithm please see Cutler & Breiman (1994) or any of the other suggested reading in the references below.

AA seeks to identify extreme/idealized versions of particular behaviors and characterize a user's behavior by its proximity to one of these archetypes. Once the archetypes have been identified, a segmentation of the observations can be obtained by assigning observations to a segment defined by the archetype they most closely associated with. Characterizing segments using extremes in this way aids interpretation of segments. However, it is important to bear in mind that the archetypes are idealized behaviors and it is not the case that every observation in a segment will exhibit such extreme behavior.

In contrast, cluster analysis seeks to identify compact sets of observations which are similar to each other but different to observations in other clusters. The cluster means/centers are typically used to characterize each cluster, rather than the extreme pure types used in AA. In the clustering setting, each observation cannot be represented as some weighted combination of the cluster means. Using AA, however, we can represent each observation as a weighted combination of the archetypes. Thus, there is a philosophical difference in these approaches.

To solidify this difference with an example, consider weather in parts of the world that have seasons. If we were to record temperature, hours of sunshine and precipitation on each day of the year and segment the days into two groups we would expect the groups to correspond to summer and winter days. An idealized summer day is usually characterized as sunny and hot with no precipitation. In contrast, an idealized winter day is usually characterized as dull and cold with some form of precipitation. The segmentation using AA would be characterized by these ideals. In contrast, the segmentation using a cluster analysis would be characterized by the average winter day and the average summer day. It is reasonable to think the weather on any day of the year can be represented as a weighted combination of the two archetypes. However, neither a very cold winter day nor a very hot summer day could be represented by some weighted combination of the cluster averages.

Other analogies often used to explain AA are the ideal physical attributes of track and field athletes for different events (e.g. sprinters are muscular and explosive while long distance runners are much lighter) or the ideal physical attributes of players in different positions on sports teams (e.g. differences between idealized point guards and idealized centers on a basketball team).

1. Cutler & Breiman (1994), Archetypal Analysis, *Technometrics*, 36 (4), 338-347.
2. Eugster & Leisch (2009), From Spider-Man to Hero - Archetypal Analysis in R, *Journal of Statistical Software*, 30 (8), 1-23.
3. Bauckhage & Thureau (2009), Making Archetypal Analysis Practical, *Joint Pattern Recognition Symposium*, 272-281.

## REFERENCES

A Guidance Framework for Developing and Implementing Vulnerability Management Gartner, 2016, <https://www.gartner.com/document/3747620>

## ACRONYMS

AA = Archetypal Analysis

KPI = Key Performance Indicator

SCA = Secure Configuration Assessment

VA = Vulnerability Assessment



7021 Columbia Gateway Drive  
Suite 500  
Columbia, MD 21046  
North America +1 (410) 872-0555

[www.tenable.com](http://www.tenable.com)



Copyright 2018 Tenable, Inc. All rights reserved. Tenable, the Tenable logo, Tenable.io, and The Cyber Exposure Company are registered trademarks of Tenable, Inc. All other products or services are trademarks of their respective owners.