WHITEPAPER

Instilling Confidence in Security and Risk Operations with Behavioral Analytics and Contextualization

Understanding Why Automated Machine Learning Behavioral Analytics with Contextualization Provides Better Results than Just Big Data Centered Approaches

Bay Dynamics®

Contents

Bay Dynamics®

Executive Summary

Big Data Analytics is a very hot topic in IT Security circles lately, specifically regarding how it can potentially be applied to derive seemingly magical results like it has in so many other applications from the Siri service on i-devices, to understanding epidemics, to predictive marketing, to scientific and engineering applications and many more. However Big Data by itself is not the panacea for all security ills that most believe it to be. This paper outlines how automated machine learning analytics with contextualization leads to better results and quickly identifies and repels attacks over approaches focused on Big Data.

The Problem: The Allure of Big Data to Security and Risk Operations and Why it is Flawed

Much of the magic in today's modern world is derived from the application of Big Data analytics concepts and technologies to combine many distinct data sets together and structure queries that answer deceptively simple sounding questions that in fact require the processing of an enormous amount of data in order to be relevant, personalized, accurate and valuable.

For example, getting a meaningful answer to the simple question "Where is the best place to eat around here?" in fact is not as simple as it might first appear.

If the answer is not fully contextualized to the one asking the question the result can not only be disappointing but potentially life threatening as well. To minimize the risks and increase the fidelity of the answer provided, the one answering the question needs to synthesize and analyze an extremely large amount of individual data

elements combined from different sources. In addition the data being analyzed must be selected and prioritized for contextual relevance including:

- Current location of the one asking
- What is meant by "around here" and what will be considered too far of a distance to the one asking using the current modes of transportation available to them
- Food preference and allergies
- Preferred price points
- And so on

The value of the answer to the one asking will be directly proportional to the amount of selectively relevant data that is brought together and contextualized as part of the analysis. Not all available data will be equally relevant or important to the decision process and not all potential data queries will lead to the desired results.

The data needs to be well understood, prioritized and rejected in accordance with the desired context otherwise it is garbage-in, garbage-out. It is this very process of selecting and prioritizing data, defining queries and lastly identifying the types of results that will be meaningful, that is the most difficult and requires the most effort in Big Data projects.

To be effective in this regard, Big Data projects need to fully establish the underlying logic that supports this decision process and this requires a deep understanding of the environment it is to operate in. This is the main flaw of Big Data projects.

Big Data's Struggle to Interpret the Raw Data

The biggest challenge in security and risk operations today is not in simply bringing together all of the security data sources. Many organizations deploy Security Incident and Event Management (SIEM) solutions or build complex Big Data warehouses that contain all of the data in a central repository. The biggest challenge in fact is providing an efficient and scalable way to analyze contextually important data in order to quickly produce a list of clearly identified and prioritized risks and threats that can be actioned immediately, well before critical incidents occur.

To achieve this goal, Big Data projects require significant expertise from expensive and hard to find data scientist who first analyze the environment, the trends, the behaviors and each of their implied meanings. To be successful, this process requires the buy-in from individual business units to help identify patterns of behaviors that indicate suspicious deviations from the norm and to build workflows that will support each scenario. Most business units struggle with identifying these behavior patterns and they fall back to an "I'll know it when I see it" response. This response results in many more weeks and months of manual data review by consultants and data scientists in the hopes of uncovering and identifying threats from each data point or data source's individual implied meaning.

It is extremely difficult, time consuming and

even impossible in many environments to code a broadly applicable "I'll know it when I see it" logic within Big Data solutions. As a result, most Big Data deployments focus on providing the Big Data system operator with advanced search and parameter combining capabilities for them to query the data to achieve this goal. While somewhat successful, this requires very experienced operators with unique domain expertise to properly use the tool and build understandings often using trial and error. This approach is un-scalable and is in fact rather risky since important patterns indicating risky behaviors may slightly be out of the scope of analysis and thus remain undetected. This is typically why attackers are able to successfully hide in the noise and remain undetected for six months or more[1].

Other Big Data deployments will attempt to address the lack of highly skilled operators by using more traditional rules-based logic akin to white-lists and black-lists. These rules are predefined with very specific sets of conditions and are applicable to only very precise sets of well-known scenarios.

The main issues with this approach are the un-flexible nature of the defined rule-set and the false sense of security it generates. Hard coded rules require significant re-coding to adapt to environmental changes such as re-organizations, new lines of business, merger and acquisitions and the deployment of new business systems. Because most rules are defined with pre-set thresholds such as "find after hours logins" or "find users accessing more than XX files or records", activities that are just below the thresholds will never surface as being important. Organizations unwittingly get a false sense of security because they do not have visibility into the actions that are just under the radar.

Lastly, even after deploying Big Data environments and SIEM solutions, many organizations continue to get overwhelmed from the onslaught of endless security log feeds. While they attempt to triage important events by leveraging severity attributions, they typically do not incorporate the impact of interdependencies between the data sources as a contributing factor in their analyses because of the amount of complexity involved.

For example a low severity event that by all appearances seems benign in one security solution can become the tipping point when taken in context with the volume, frequency, severity and velocity of events in another security solution. This inter-relationship becomes indicative of broad risk trends and is the only way most organizations will ever have to identify a breach that is in progress. What happens in one security solution must influence the meaning of the others. Again, in this scenario, Big Data environments must be manually configured to not only understand individual events but also the interdependencies and their implied meanings. For most organizations the limitations of finding

[1] M-Trends 2015: A view From the Front Lines

experienced staff to manually code rules, project budgets and time make this approach a practical impossibility.

How Automated Machine Learning Behavioral Analytics with Built-In Contextualization Correct the Flaws of Big Data Approaches

An Automated Machine Learning User Behavior Analytics with built-in contextualization and predictive analytics capabilities solution is designed out of the box to address the issues impacting traditional static analysis approaches found in Big Data and SIEM deployments.

As a key starting point, Automated Machine Learning Analytics with built-in contextualization do not require any manual tuning of rules or the understanding of the underlying implied meaning or importance of the data sources or their dependencies.

It provides this level of capability out of the box by leveraging a "let the data tell the story" approach. The logic is pre-built to quickly identify deviations from the norm rather than identifying specific hard coded pre-established behaviors that are subject to under the radar attacks. It does not pre-dictate what the data should be saying but rather lets the patterns of behaviors speak for themselves. This means it is fully in line with the preferred business-friendly "I'll know it when I see it" approach where anomalous behaviors are automatically identified and prioritized for review and automated workflows.

The solution automatically identifies the important elements that matter the most from your security data feeds and automatically connects the dots across your organizational assets. In this way, the system is aware of dependencies and takes into account the impact of events in individual security solutions together in context.

The solution also establishes fully adaptive activity baselines for all the entities within your environment (users, administrative accounts, system accounts, applications, etc.). This approach automatically incorporates changes within the environment over time as they occur without the need to manually code them.

By focusing on entities instead of individual events, these solutions directly align with the way organizations naturally operate and also provides for significant leverage in investigations. Each user can generate thousands of events every day. Correlating events to users and then identifying when they deviate from the norm effectively reduces the scope of the work from hundreds of thousands of events to a few users that have a high likelihood of being at risk. Organizations can scale operations to support the investigations of a few users but they cannot scale to reviewing hundreds of thousands of events.

The other advantage to a "let the data tell the story" approach is that it eliminates the threat of "seeing what you want to see, not what you should see" analytics. If an organization has never seen an attack pattern previously, how can it

code a rule to identify it? How can an organization create a query that will find a behavioral pattern that is anomalous in a way not previously documented without a lot of blind luck? It cannot. As a result most organizations create rules coded from patterns that are familiar, that have been seen before, until a new pattern is identified and then coded into the system. The drawback from hard coded approaches is that organizations are perpetually behind the attackers who constantly evolve their attack approaches. This is why it is important to "let the data tell the story" and let it "tell you what you should see, not just what you want to see". This approach is the only way for security and risk operations to stay relevant in today's highly dynamic threat environment.

Lastly the solution automatically and continuously monitors for deviations from the norm. Deviations are automatically prioritized in terms of severity and can be automatically assigned to remediation paths that range from reviews by line of business managers, HR, Security Operations to Just-in-Training sessions for non-malicious policy violations to triggering investigative actions.

Benefits of Automated Machine Learning Behavioral Analytics with Built-In Contextualization over Big Data Approaches

Because of the inherent flexibility Automated Machine Learning Behavioral Analytics with Contextualization solutions provide, findings can drive new forms of personalized and directly actionable security and risk insights that make it possible for individual business-side stakeholders to become a key part, as they always should have been, of their own portion of the security and risk management program without requiring them to directly interface with technical solutions or be overwhelmed with the raw data. This is an extremely powerful evolution in security and risk management programs that ties-in direct accountability with the actual business acceptance of risk.

Organizations can also "do more with less" because the solution is designed to operate out of the box without the need for manual configurations or consulting to establish acceptable baselines or tuning to reduce false positive or create alerts for false negatives. Leverage is applied when focusing on users instead of events and fewer resources can effectively manage the same workload because the system automatically prioritizes the most important and risky behaviors.

When organizations deploy solutions that focus on the business risk impacts of IT environments and provides them with clearly actionable insights that cuts through the noise, and eliminates false positives and false negatives, they can properly assess the effectiveness of day-to-day security and risk operations in full context. This means that organizations now have the level of confidence that their own security and risk operations are in fact addressing all of their risks associated with the insider, outsider, high privilege user and the attack surface of their most sensitive assets.

Introducing Risk Fabric: A Predictive Security Analytics Platform

The evolution of cybersecurity is driven by a new breed of modern threats and is fueled by organizations inability to effectively detect and respond to these threats. The anatomy of modern threats goes beyond legacy and advanced malware; it comprises of users, attackers, applications, high privilege users and high value assets.

Bay Dynamics is committed to protecting organizations against the modern threats and reducing their business risk with its innovative next gen machine learning and predictive analytics approach. Using this approach, Bay Dynamics leads the paradigm of continuous "threat detection with proactive response" for organizations. Bay Dynamics' Risk Fabric Platform powered by the Predictive Security Intelligence™ (PSI) Engine harnesses this innovative approach to offer four business solutions namely insider threat solution, outsider threat solution, attack surface threat solution and high privilege access threat solution. These four "out of box" solutions deliver complete threat protection and actionable intelligence by detecting and contextualizing anomalous user behavior (i.e. attacker masquerading as insider or malicious insider); continuous monitoring of high privilege access across users and systems and proactively identifying exploitable high value assets.

About Bay Dynamics®

Bay Dynamics is the market leader in providing cybersecurity solutions to protect organizations around the world from evolving security threats and to reduce their business risk with its innovative next generation machine learning and predictive analytics technology. For more information, visit www.baydynamics.com.