WHITE PAPER | MARCH 2017

Enterprise Data Security: The Basics of User Behavior Analytics

ca technologies

Table of Contents

Executive Summary

Reports of cyberattacks now dominate the headlines. And while most high-profile attacks—including the major breaches at JP Morgan, Anthem and Slack—originated outside of the victimized organizations, theft and misuse of data by privileged users is on the rise.

In fact, 69% of enterprise security professionals said they have experienced the theft or corruption of company information at the hands of trusted insiders.[1] There are also cases where a company's third-party contractors, vendors or partners have been responsible for network breaches, either through malicious or inadvertent behavior.

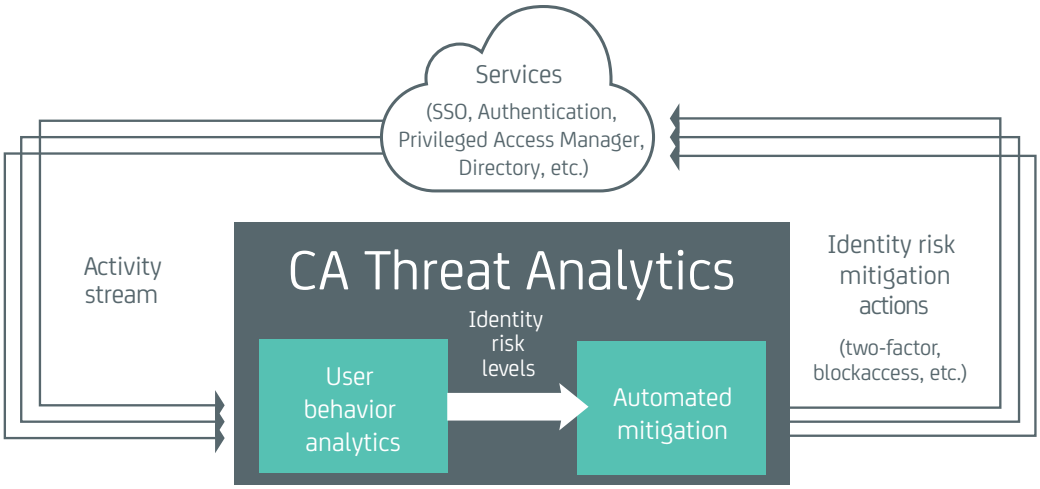
If events such as these have taught us anything, it's that protecting privileged access remains an urgent concern for companies of all sizes. But despite this awareness and a surplus of available security products, many IT systems are still vulnerable to attack.

The fact is, traditional identity and access management (IAM) controls, though extensive, are static. And once a malicious user gains access, they are free to exploit the system up to the extent of the account's set privileges.

But by deploying an identity-centric approach to security that brings together user behavioral analytics and anomaly detection into a self-learning model, enterprises can quickly detect risky activity and automatically trigger mitigating controls to limit damage to the enterprise.

CA Threat Analytics

CA Threat Analytics protects enterprise data the same way that credit cards protect money. While this phrase evokes the right ideas—persistent monitoring, the use of analytics to determine risk and prevent the 'bad guys' from stealing assets—it does not offer much insight into how this is accomplished. This white paper outlines how CA Threat Analytics protects enterprise data using two related capabilities: user behavior analytics and automated mitigation.



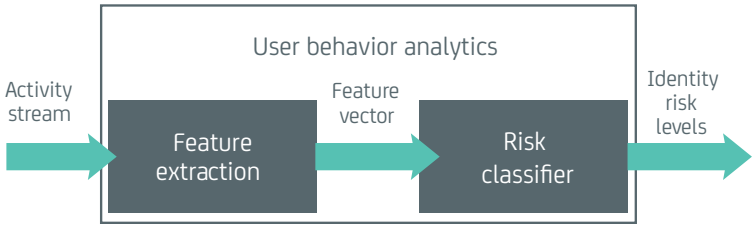
User behavior analytics enables the enterprise to continuously assess risk and quickly detect malicious activity. As input, user behavior analytics takes a stream of data about how a given identity or group of identities interacts with services or applications—then outputs a level of risk associated with each enterprise identity.

Automated mitigation enables the enterprise to automatically take steps that mitigate risk and thwart detected malicious activities. Automated mitigation changes how access is controlled for individual identities based on the risk output of the user behavior analytics. A simple example of an automated mitigation would be to automatically block a high-risk identity's access to a particularly sensitive app or data repository.

Although both user behavior analytics and automated mitigation are integral to how CA Threat Analytics works, this white paper purposefully focuses on user behavior analytics. In the following sections, the user behavior analytics function pictured above will be broken down into its constituent parts. Then, these parts will be discussed individually in detail. For the sake of simplicity, the discussion initially centers around protecting a single identity on a single service. After explaining the basics of the techniques used, there is discussion of how these ideas can be enhanced when working with a population of identities across multiple services.

The Basics

Conceptually, the user behavior analytics function is made up of two components: feature extraction and the risk classifier.



The feature extraction component processes an activity stream and extracts a set of relevant features. The relevant features are characteristics about an individual identity that have been observed over time, such as:

- The identity is using an unknown mobile device.
- The identity is operating in a remote location.
- The identity is coming from a suspicious IP address.
- The identity is a member of a privileged group.
- The identity used service X outside of their normal operating time.

Feature extraction is more complicated than it appears because it's not simply extracting characteristics about an ongoing transaction. Although an activity stream arrives as a sequence of discrete events, the real input is the complete activity stream from the beginning of time. This allows you to understand aggregate usage and behavior about each identity. Without examining the full activity history, you would be forced to evaluate risk based solely on each discrete event individually.

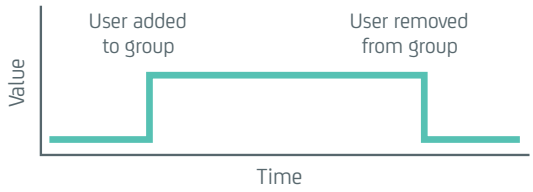
Using an example from the features listed, what does normal operating time mean in the context of a single event? For CA Threat Analytics to be able to use important features like this, it needs to calculate and use insight regarding historical data as well.

By examining the full activity stream, CA Threat Analytics provides the enterprise considerably more insight than has previously been available to assess risk and detect malicious activity. The enterprise can now assess risk based on past activities and specific information about individual identities. This benefit comes at the cost of processing a large amount of data, much of which is redundant. Fortunately, by performing feature extraction, the dimensionality of the data is reduced. This eliminates or aggregates redundant data while highlighting the information needed by the second part of the user behavior analytics function: the risk classifier.

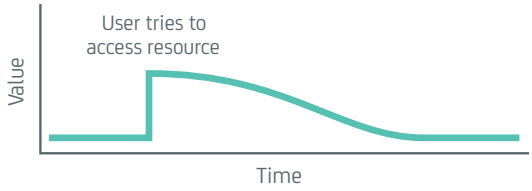
Determining Value in the Context of Time

Before we move on, let's point out an interesting detail of the features being observed over time. Because they are modified when activities arrive, they technically live in the time domain; this simply means that the values change over time. When a feature is observed, CA Threat Analytics models the observation as a function of time. In other words, if an incoming activity causes a feature to activate, the feature's 'value' may be at a maximum at the time of that activity and change as time moves forward.

The actual way the value changes varies widely depending on the feature that has been extracted. Some are fully binary, so when the feature is observed it stays at its highest value until something mitigates it, as below.



An example would be membership in a sensitive group. That feature is full-valued for the entire span of time the identity is associated with the group. Other features are modeled as decaying pulses. When a feature of this type is observed, the value is highest and it decays over time, as below.



An example would be when a user attempts to access a resource for which they do not have permission. Though that feature is relevant to an identity's risk level today, it is much less relevant in a week and even less so in a month. By decaying the value of the features over time, CA Threat Analytics ensures that the features contribute to risk in the most relevant way.

The Risk Classifier

The risk classifier is an analytic function that converts the feature vector into three discrete levels of risk:

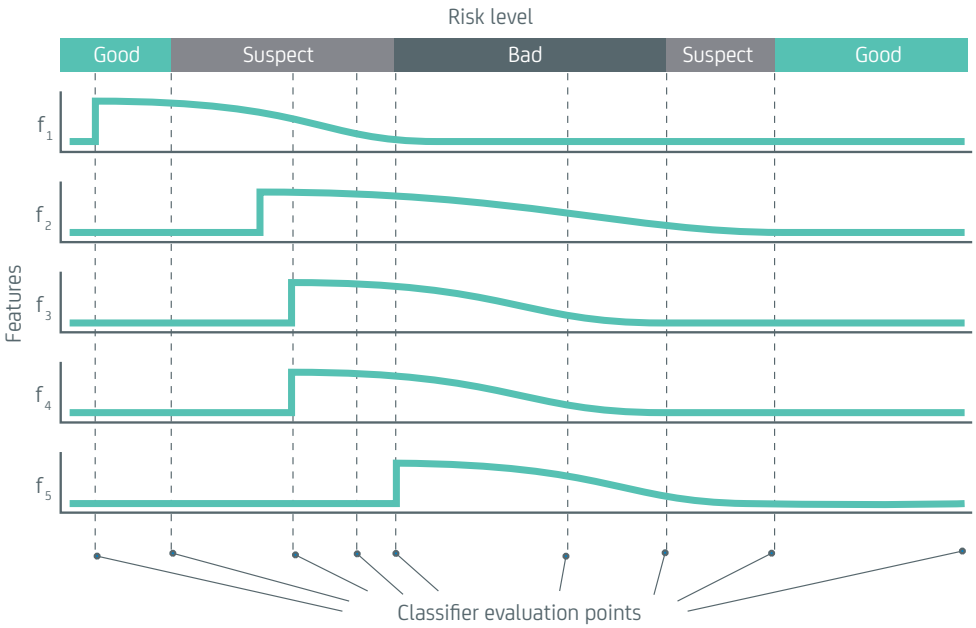
- **Good**—The identity poses minimal risk.
- **Suspect**—The identity has been associated with events or activities that pose risk, but this risk does not demand immediate action. The system will monitor this identity more closely and may initiate an initial set of automated mitigation, depending on the enterprise's policy.
- **Bad**—The identity is considered a high risk and merits immediate attention. The system will initiate automated mitigation and alerts, per the enterprise's policy.

The risk classifier functions take as input a vector of feature values and outputs one of the discrete classes above.

$$\text{Classifier}(\text{features}(t)) \longmapsto \{\text{good, suspect, bad}\}$$

As discussed above, the features themselves are a function of time, so the risk classifier function also operates in the time domain. The risk classifier is invoked at critical decision points, typically in response to significant changes in the values of the feature vector. Whenever the risk level is calculated by the risk classifier for a given point in time, all of the feature functions are evaluated for that identity or entity at that moment. The full set of features active for the entity at that moment make up the actual feature vector consumed by the risk classifier and is used to determine risk.

In the figure below, the discrete points at which the risk classifier would likely be evaluated are annotated. As noted, the evaluations occur when a feature increases in value and when a feature value falls below a threshold. The values passed to the risk classifier correspond to the value of each feature at the point in time when its evaluation is triggered, corresponding to the vertical lines above. Of course, not every run of the risk classifier results in a new risk level. Practically speaking, there are many more evaluation points than pictured, corresponding to changes in feature value, system activity, threat intelligence. In general, the risk classifier is activated any time that there might be a change in risk level.



So, what is the risk classifier itself? How does it translate a feature vector into one of a discrete set of risk classes? It helps to start by stating what it is not. CA Threat Analytics risk classifiers are not simple rules that test for specific features—such as "if feature X is active, return bad." This is a naïve approach that's used by many traditional security products. This approach fails spectacularly because it's highly prone to false positives, brittle and easily defeated. Additionally, it does not make use of the information that's critical to both detecting malicious activity and making the system usable for legitimate users.

CA Threat Analytics capabilities are far more robust. The CA Threat Analytics risk classifier examines features not in a vacuum, but in the context of the entire feature set. With this approach, several features—that in isolation have no impact on risk level—can combine to affect risk in a meaningful way. What's more, CA Threat Analytics incorporates feedback from deployed systems, including aspects of individual users and changes in the population of identities, to refine the decisions it makes over time. The result is a system that provides the flexibility to adapt as new threats and deployment scenarios emerge.

Populations and Services

As mentioned earlier, there are several practical details that have been simplified for the discussion above. First, what about populations of identities? Especially in the enterprise environment, there are aspects of the group of identities that are relevant to the risk level for a given identity. A few examples:

- Accessing resources with more devices than is normal for the organization
- Operating outside of the group's normal operating location
- Being in an inappropriately large number of groups

For each organization, the baselines of expected activities—which include factors such as the normal number of devices associated with a user, the organizational operating locations, the appropriate number of groups—are different. By looking at a group of identities rather than identities in isolation, you can get a high level of useful population statistics against which you can compare the individual identities. Of course, this comes at a cost. Instead of processing merely the entire activity stream for an identity, this requires performing feature extraction on the full activity history for the entire organization.

In a similar way, extending analysis from a single service to a group of services offers a different benefit. By examining the actions an identity takes across different services, we can extract features to build models of typical access patterns and intelligently apply them to provide security across the services. This information enables CA Threat Analytics to detect anomalous and inconsistent behavior that represents a threat to that identity or the enterprise.

Conclusion

This white paper introduced how CA Threat Analytics protects enterprise data using user behavior analytics. While the basic ideas are easy to explain, the practical issues with feature extraction and risk classification are well beyond the scope of this paper. Indeed, many of the real-world requirements that drive our team were not even mentioned—such as enabling real-time decision making, ensuring system accuracy over time and providing system admins with true insight regarding risk decisions.

If you're interested in finding out more about these drivers and how your organization can benefit from them, please see: https://www.ca.com/us/products/ca-threat-analytics-for-privileged-access-manager.html

Connect with CA Technologies at ca.com

CA Technologies (NASDAQ: CA) creates software that fuels transformation for companies and enables them to seize the opportunities of the application economy. Software is at the heart of every business, in every industry. From planning to development to management and security, CA is working with companies worldwide to change the way we live, transact and communicate – across mobile, private and public cloud, distributed and mainframe environments. Learn more at **ca.com**.

1 Accenture and HfS Research, "The State of Cyber Security and Digital Trust 2016," June 2016: https://www.accenture.com/t20160704T014005__w__/us-en/_acnmedia/PDF-23/Accenture-State-Cybersecurity-and-Digital-Trust-2016-Report-June.pdf#zoom=50

CS200-259092