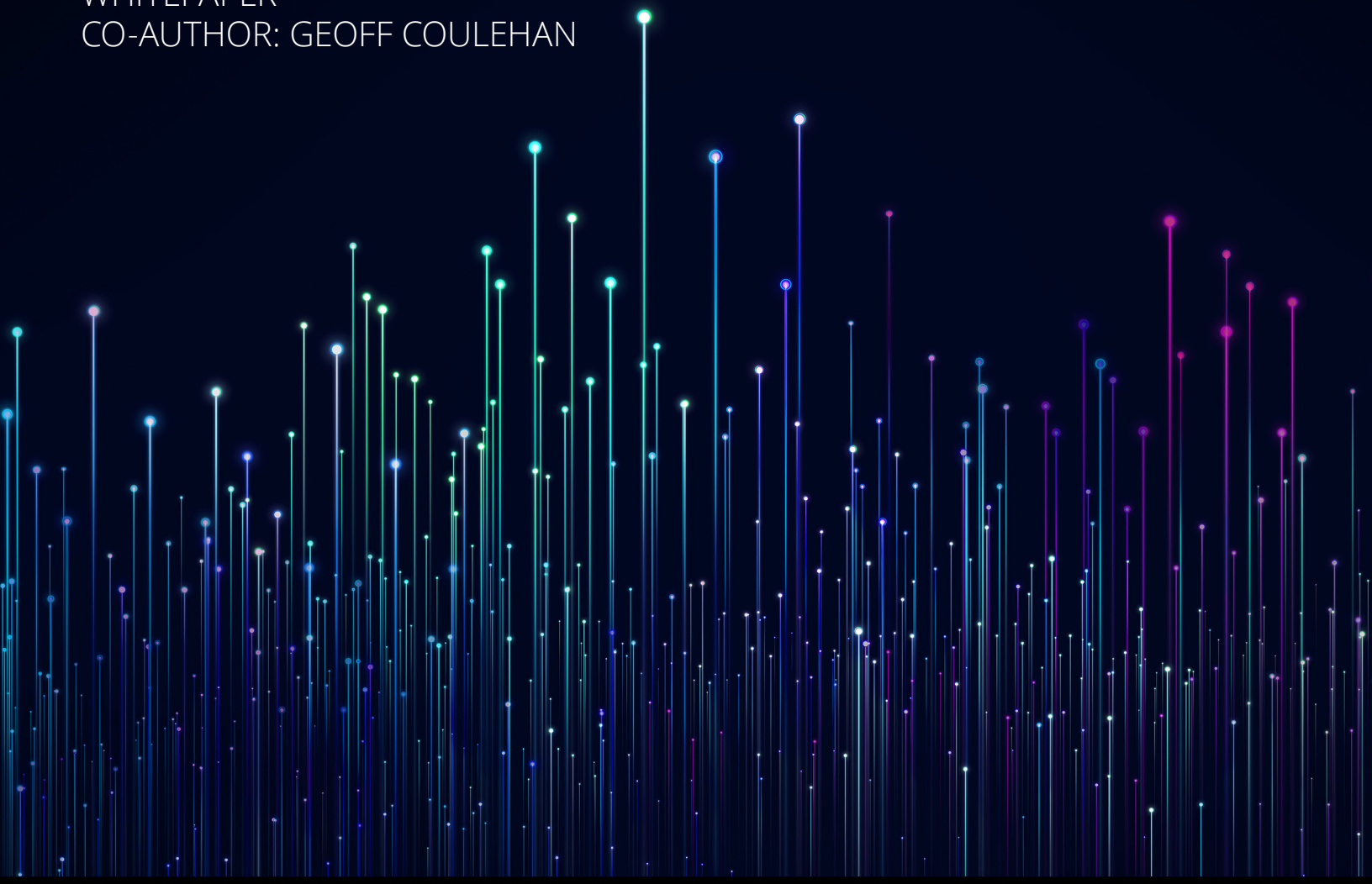
MixMode™

HOW PREDICTIVE AI IS DISRUPTING THE CYBERSECURITY INDUSTRY

WHITEPAPER

CO-AUTHOR: GEOFF COULEHAN



Contents

"With MixMode we're flipping the entire model on its head, putting AI First, in front of flawed and costly data aggregation processes"

Key Takeaways

- Most cybersecurity solutions that claim to use "AI" are touting manual, rules-based technology that requires security professionals to continuously extract, normalize, and analyze exclusively historical data before "AI" initiates.
- An evolving or "generative" baseline to understand "normal" network behavior coupled with an self-supervised, AI first comparison of current network conditions is the only way to do effective anomaly detection.
- Network baselines are useless when based on historical data. Building and understanding a true network baseline which evolves and adapts to network conditions over time requires true self-supervised "Third-Wave AI."
- Traditional SIEM approaches to cybersecurity are ineffective and "additive" in terms of overall cost, infrastructure, and human labor, and they contribute to an unnecessary increase in data migration, redundancy and latency. Adding an AI layer with MixMode is complementary to a SIEM and increases productivity and efficiency.
- Zero-day security threats require modern solutions. Rules-based security tools are fundamentally flawed and insufficient against bad actors who understand how these platforms work.
- Solutions like MixMode which utilize "Third-Wave" Context-Aware AI are the only way to create an evolving, accurate baseline and are orders of magnitude more cost-effective, resource-effective, and far less infrastructure intensive.

Introduction

As cybersecurity evolves and bad actors become more sophisticated, organizations must also evolve. Security teams must take a more proactive approach to Network Traffic Analysis (NTA) in order to avoid the next generation of hacks and breaches to ensure a sound cybersecurity posture. Standard industry solutions include so-called artificial intelligence models that are fundamentally flawed in that they compare network behavior exclusively against a historical baseline analysis that is created after months of data is aggregated, stored, and analyzed.

Having an accurate, forward-looking, and evolving baseline of "normal" network behavior to measure anomalous activity against is the only reliable and accurate approach when fighting against a slew of new bad actors and attacks. However, a major problem exists for cybersecurity solutions that claim to deliver anomaly detection through AI: the baseline they create and measure against is based exclusively on historical data which takes months to gather, creates ever-increasing false positives, and does not support anomaly detection as network conditions and attackers evolve.

Without an accurate, generative baseline that evolves over time, truly meaningful anomaly detection is impossible.

In contrast to many cybersecurity solutions which are based on Supervised Learning or "second-Wave AI" which requires constant training, human tuning and historical data, **"third-wave AI"** solutions (as defined by DARPA), which leverage generative, self-supervised learning, can offer an accurate evolving baseline of normal network behavior in real time and predict appropriate future network behavior. This approach allows MixMode to provide extremely accurate anomaly and threat detection, 95% fewer false positives, and predictive threat detection.

This paper will evaluate several common SecOps issues around Network Traffic Analysis, explaining why typical solutions are wholly ineffective and represent sunk costs versus added value. We'll examine how self-supervised learning AI is poised to overcome the SecOps challenges of protecting today's distributed networks.

We'll examine the current state of the cybersecurity solutions marketplace:

1. **Moving Beyond Rules-based AI Solutions, Making Sense of the AI-Enhanced Cybersecurity Market**

2. **The Inherent Inefficiency and Inaccuracy of Stand-Alone SIEM Platforms**

We'll take a look at three security operations center issues negatively impacting Network Traffic Analysis:

1. **The Wasteful Culture of False Positives and the Wasted Potential of Security Analysts**

2. **The Human Error Factor**

3. **The Shifting Definition of "Baseline"**

We'll consider current research and statistics that help to shape the story of what's happening in the security platform stratosphere, and share insights from Geoff Coulehan, Head of Strategic Alliances at MixMode about game-changing, third-wave AI in Network Traffic Analysis and cybersecurity.

In addition to serving as the Head of Sales and Strategic Alliances for MixMode, Coulehan has honed his industry expertise over two decades spent examining and improving the continually evolving cybersecurity landscape.

Moving Beyond First- and Second-Wave AI Solutions

Solutions using first and second-wave AI are far less effective than third-wave context aware solutions. To understand the limitations of early AI-enhanced security solutions, we must consider why engineers developed these AI functions in the first place.

First-Wave AI

First-wave AI adds automation to repetitive, narrowly defined tasks. For example, tax software that operates based on predefined rules and past behaviors, but can't perform functions beyond these limitations.

First-wave AI is purpose-built and intended to solve specific problems. It addresses security challenges based solely on human inputs related to that problem. While the introduction of first-wave AI was a remarkable leap in the world of technology, it is clear that first-wave AI security capabilities are insufficient across sprawling, distributed networks.

Second-Wave AI

Second-wave AI is more nuanced in its classification and prediction capabilities but has only minimal reasoning capability. For example, IBM's Watson can process data to provide insight and answers but is not able to understand the context or explain how it figures out a solution.

SIEM cybersecurity platforms represent a common modern usage for second-wave AI. While this software usually includes automated functionality and some unsupervised behavior, it still requires a great deal of ongoing human interaction, tuning, configuration and guidance.

Third-Wave AI

Third-wave AI platforms, like MixMode, are context-aware, and use a generative model based on self-supervised learning to go beyond unusual activity identification and can predict future outcomes.

"MixMode starts learning from the first five minutes it is deployed, does not require historical data, and is adapting actively to the dynamic changes in massive amounts of network data."

Industry AI and Automation Analyst

Making Sense of the AI-Enhanced Cybersecurity Market

The cybersecurity and network traffic analytics sectors have become oversaturated with vendors who make claims about AI that are often overblown and patently false. For example, some vendors refer to scheduled reports as AI. When vendors refer to basic, automated, and semi-automated functions as AI, it's difficult for SecOps teams to differentiate available security tools.

The typical "AI solution" offered by cybersecurity tool vendors requires customers to extract, normalize, and analyze historical data before "AI" initiates. This process is costly, arduous, tedious, takes months to perform, and leverages data that is, by nature, "out of date."

To properly assess network behavior as it occurs, **AI must happen first.**

"With MixMode," Coulehan says, "we're flipping the entire model on its head, putting AI First, in front of flawed and costly data aggregation processes."

Third-wave AI, such as MixMode, is self-supervised, generative, and predictive. When the MixMode platform establishes a baseline, it doesn't need to get up to speed by analyzing months or years of data. Recently IDC AI Analyst Ritu Jyoti explained why MixMode is different from other solutions in this regard: "MixMode starts learning from the first five minutes it is deployed, does not require historical data, and is adapting actively to the dynamic changes in massive amounts of network data." The advent of Third-wave AI in MixMode is ultimately what makes it effective enough to identify anomalous behavior on a network before the information has been aggregated, with little human intervention.

An AI solution is not comprehensive if it can't automatically adjust to shifting baselines in real-time. When real-world events impact network behavior, typical security products go into hyperdrive, generating false positives.

The global response to the Coronavirus pandemic, for example, has impacted typical network behavior for countless organizations. A significant portion of workers (up to 80% from 20% almost overnight) who rely on network connectivity have either been laid off or have suddenly shifted to a home-based work environment. Coulehan points to behaviors like a decrease in file access across shared drives and an increase in the usage of applications requiring more bandwidth and alternative data access methods.

The result of this atypical but innocuous behavior, and utilizing AI that is based on historical data (in which a pandemic was not sweeping across the globe) means that the behavior would increase in false positives triggered by anomalous, non-threatening behavior. Without adequate machine-learning capabilities, security platforms can't adjust.

Coulehan says that without a clear understanding of what "normal" looks like on a given network at a given time, anomaly detection and prioritization is impossible, by definition. Because MixMode utilizes third-wave, generative AI, tuning itself over time to predict what normal should look like moving forward. The platform surfaces anomalies, as well as threats before data has been aggregated, stored, and optimized.

"Vendors in this space typically focus on identification of events or indicators as a starting point for threat intelligence processing," Coulehan says. "What they intentionally avoid is this idea of real-time anomalies."

The Inherent Inefficiency and Inaccuracy of Stand-Alone SIEM Platforms

While there are a myriad of cybersecurity solutions available, the overwhelming trend has been to choose a standalone SIEM. These popular first and second-wave AI software solutions can seem ideal.

The SIEM or Security Information and Events Management process makes sense at a high level. SIEM, true to its name, are fundamentally search and investigation platforms focused on historical data analysis. As we've discussed, however, the nature of SIEM results in an endless flow of ever-increasing false positives that require skilled security analysts to devote valuable time to threat hunting.

SIEM software operates via data logging functions based on information gathered from a variety of system resources. Common data sources include:

- Firewalls
- Antivirus Filters
- Host Systems
- Applications

The SIEM platform flags every potential security incident or event as it analyzes the collected data. High-risk events trigger an alarm so that SecOps teams can respond as they occur.

The Additive Nature of Common Cybersecurity Solutions

Coulehan considers these traditional SIEM approaches to cybersecurity "additive." "It's additive in terms of overall cost, infrastructure, contributes to an unnecessary increase in data migration, redundancy and latency," he explains. "Second-wave AI doesn't address any of those problems, even if it's self-supervised."

"Typically, cybersecurity solutions have required clients to collect information from all of their peripheral network components," he says. "To extract, normalize, and store log files of all their next-generation firewalls and machine data that is providing insights into the network." The log files are then consolidated into a single location and normalized for field mappings so the organization can develop a single system of truth.

Problems develop when organizations need to add new data points to feed into their SIEM or security platform. Coulehan says false positives go up, as do expenses. "Manual intervention of skilled personnel required to tune, operate, and run the system increases linearly and in parallel with aggregate data requirements, false positives, and alerts," he adds.

It's no secret that security vendors rely on expanding data stores to increase profit. Common SIEM pricing structures include agreements that charge in one of three ways:

- Data volume measured in events per second
- Data indexed
- Volume of average data being processed

Gartner estimates that enterprise data volume is doubling on an annual basis. More data means expanded license costs; organizational growth requires more data. Coulehan says the additive business model leads directly to less secure environments. "To save on costs, organizations often eliminate or segment sources of data that typically should provide them with the most value, insights, and the best set of information to identify anomalous behavior," he explains. "They exclude the most valuable data day one because it's just too expensive." "If you're trying to keep the cost for a system down by eliminating the most useful source of information, the approach is fundamentally flawed," he adds.

Other Product Limitations Unique to SIEM

High Touch Environments and Human Input
These systems require a great deal of human input and analysis. Even when the SIEM functions well, analysts will spend tedious hours on manual review. As previously discussed, human error has a very real and expected impact on security outcomes.

Accuracy Issues

SIEM platforms base their analyses on only the amount of information they can log, and they can't log everything. SIEMs miss security threats routinely, including highly damaging threats.

Logged Data Risks

Hackers are aware that SIEM relies on logged data to assess risk. It's no surprise that SIEM-created logs have become a favored target for bad actors, who gain access to steal and destroy data while tampering with logs to cover their tracks or mislead analysts.

Talent Shortage

To use SIEM effectively, talent matters. Security analysis is not a run-of-the-mill tech job. "Unfortunately," Coulehan adds, "it's a costly position dependent on highly skilled individuals, tasked with repetitive, mundane alert, search, and investigative tasks, plagued by high turn over." There's little point in investing in SIEM software if an organization cannot attract and retain talented professionals to run it. There will be 3.5 million unfilled cybersecurity jobs by 2021, according to the Herjavec Group's Managed Security Services Provider (MSSP) 2019/2020 Annual Cybersecurity Jobs Reports.

Expense

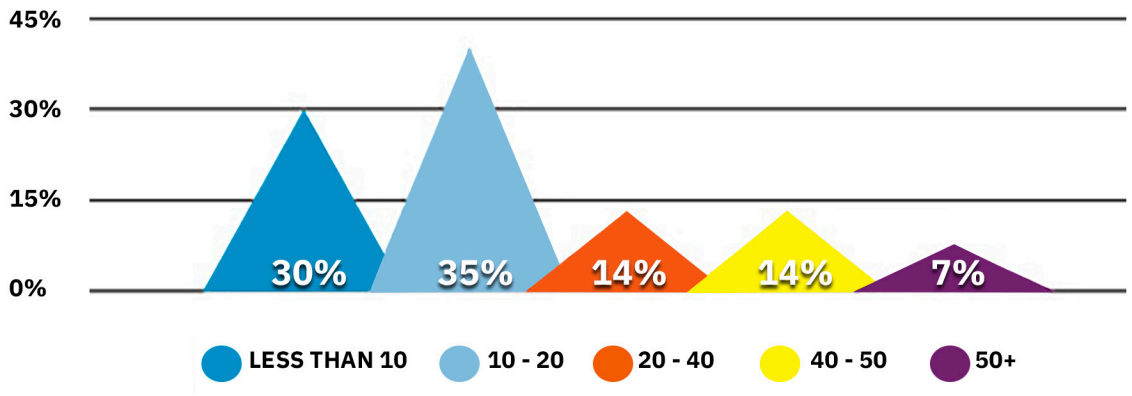
SIEM is expensive. Large enterprises invest hundreds of thousands of dollars or millions in SIEM software, hardware, and associated personnel cost. The cost doesn't stop adding up even once the system integrates into an organization's network. As Coulehan points out, this type of security solution fundamentally relies on a growing mass of data that needs to be analyzed and protected.

"SIEM solutions require ever increasing volumes of data that they ingest," he explains, "so the traditional approach for vendors is to continually increase the volume of historical data they're capturing and storing. Then, they will add targeted solutions behind that aggregate data."

Regulatory Compliance Challenges

Regulations like the GDPR and CCPA require all organizations that handle consumer data to protect, track, and provide responsive answers to consumer data queries. SIEM software has to be configured manually and ideally, mapped directly to regulatory requirements. This is how it can identify security incidents and prove that an organization has the required security controls in place. It is expensive and time-consuming to add compliance functionality to SIEM software.

How many incidents/alerts do you personally investigate per day on average?



Source: CRITICALSTART SOC Survey

Dynamic Data Challenges
A security solution is only as reliable as the data it can analyze. When that data is dynamic, as is the case with the majority of data collected and used by organizations today, SIEM is only as current as the latest update. This is an especially challenging hurdle for increasingly-popular bring-your-own-device (BYOD) and Internet-of-Things devices linked to enterprise networks. AI solutions that don't include machine learning capabilities can't keep up.

Three Security Operations Center Issues Impacting Network Traffic Analysis

1. The Wasteful Culture of False Positives and the Wasted Potential of Security Analysts

Coulehan points to the massive amount of false positive alerts triggered by SIEM platforms as a major issue for organizations on several fronts. Not only are they wasteful of time and human resources, they ultimately open up organizations to more vulnerabilities and lead to a cascading series of security events. The opportunity cost of directing resources to alerts that turn out to be false positives can be a serious network breach that happens in the meantime.

A recent Ponemon Institute report revealed that a typical organization wastes between 286 and 424 hours every week handling false positives. Time wasted hunting these threats is one of the top issues contributing to ineffective security operations centers (SOCs), according to the report. Ponemon reported that 49 percent of companies say false positives are a top challenge.

How Threat Hunting Is Changing the Role of Security Analysts
Organizations historically hire security analysts to protect company assets and ensure regulatory compliance. Both of these primary goals are at risk when analysts are tasked with accounting for every false positive they encounter. The result is an endless cycle of incomplete threat management.

Coulehan says security analyst duties have become incredibly difficult and stressful. "False positives exponentially compound the difficulty of their job function because they have to comb through every threat," he explains. "They have to demonstrate to their management team that they are examining, prioritizing and resolving a never ending stream of alerts, resulting in "Alert Fatigue."

Further, Coulehan says many organizations actually consider their investment in security analyst teams as a sunk cost.

It's hard to argue with that perspective when we consider some alarming stats around the issue of false positive hunting:

- Nearly half of SecOps teams encounter false positive rates of 50 percent or higher from their security platforms. (2019 CRISTICALSTART Impact of Survey Alert Overload Study)
- As much as 25 percent of a security analyst's time is spent chasing false positives—every hour an analyst spends on the job includes 15 minutes wasted to fruitless threat hunting. (2019 Ponemon Institute Research: Improving the Effectiveness of the SOC)
- Eight in 10 SecOps teams experienced high turnover in 2019—two in 10 reported more than 40 percent analyst churn. (2019 Ponemon)
- Enterprises spend $1.3 million and waste 21,000 hours annually dealing with false positives (2019 Ponemon)
- 38 percent of SOCs report being understaffed. (The Exabeam 2019 State of the SOC Report)

The reality is that security analysts are highly-skilled assets. Organizations waste this talent when available security tools are insufficient for managing modern security threats. The human brain is capable of high-order analysis. Still, threat detection and coordinated response must happen as it occurs, which is near impossible without automation and real machine learning AI. As Coulehan puts it, "It's a high-stress job, one that is heavily redundant. Humans, by our very nature, are not well suited for highly repetitive, complex tasks."

"MixMode can help analysts provide transparency to their leadership team about what they're spending their time focusing on and why"

How MixMode Improves the Lives of Security Analysts

The good news is that the advent of third-wave AI is ushering in an era where security analysts can apply their talents in more productive and more profitable ways. Because the MixMode platform can decrease noise in security systems by 90 percent in a single week, SecOps teams can focus on tasks that improve organizational effectiveness, free from the burden of excessive threat analysis.

The MixMode platform is API-driven and can make the data sent to a SIEM, SOAR or other platform more precise and more accurate.

"Mixmode can help analysts provide transparency to their leadership team about what they're spending their time focusing on and why," Coulehan says. "It makes their job more interesting, their work more effective, and it better protects the entire organization, simply by doing what we do out of the box, free from human intervention."

MixMode's generative AI turns the typical security model on its head. As the platform integrates into an organization's network, it begins to learn what a normal baseline should look like and what anomalous behavior looks like, and continues to optimize, and update the baseline without human intervention.

2. The Human Error Factor

Because analysts are human, SOC and CISO managers expect a certain rate of human error throughout security processes. Research firm Gartner estimates that by 2025, more than 85 percent of successful attacks against modern enterprise user endpoints will exploit configuration and user errors.

In a public report in 2018, Dave Hogue, technical director of the US National Security Agency's (NSA) Cybersecurity Threat Operations Center, stressed the impact of human error on NSA security incidents over the previous year. Hogue attributed some 90 percent of NSA incidents to human error.

A recent Kaspersky Lab report cites a similar statistic around cloud-based data breaches. Kaspersky reported that employees themselves were to blame for 90 percent of these breaches, at a corporate cost of $1.25 million to $8.19 million each.

Cybersecurity Processes Vulnerable to Human Error

Effectively, any time humans interact with technology, there is a risk of error. When it comes to security, human error can correlate directly with network vulnerability from a few angles:

- Incorrect or outdated security platform configurations
- Inaccurate analysis of SIEM threat alarms (missing true positives)
- Misconfigured access controls across the network

While SIEM performance is undoubtedly more useful than the analysis a SecOps team could produce manually, it is still negatively impacted by the human factor. Considering the shortage of available qualified security analysts, this is a troubling aspect of risk mitigation.

The relationship between SIEM and human interaction is inextricable, yet this basic fact lessens its effectiveness. A SIEM platform is limited to the data stored on a system, and that data has human fingerprints all over it. Humans set up configurations for firewalls, set access controls, and ultimately decide which threats are severe enough to address, all based on network data.

In truth, the behavior of a network in response to human interaction gives us much richer, accurate insight into risk assessment.

3. The Shifting Definition of "Baseline"

Remember that a baseline is an examination of a network's performance in real-time. On the face of it, a baseline is a relatively simple concept. When it comes to cybersecurity platform solutions, however, vendors tend to gloss over the details. Would-be clients are wise to ask critical questions of a vendor:

- How will you capture the baseline?
- What data points will the baseline include and exclude?
- Does the platform actually use baseline data in real-time, or do analyses rely solely on historical data?

Traditionally, security companies sell service plans aimed at generating recurring annual revenue. Often, vendor agreements include licensing requirements that create a recurring revenue stream.

Coulehan says that this model is limited in scope because of baseline-related issues inherent to these limited solutions. "The reality of the situation is, in order to understand and establish even an exclusively historical baseline, it takes a lot of time, a lot of effort, and a lot of data,and a lot of money. By the time you actually go through this exercise of aggregating and understanding the data, it's out of date. Compounding the problem, the baseline will inevitably change dramatically, even as normal behaviors change," he says.

Without an understanding of what a normal baseline looks like or even what data it should include, organizations are missing a critical component of any cybersecurity threat intelligence or network analysis project they undertake.

Typical security solution vendors tend to skip right over this crucial element. The "baseline" some service providers create adds little value when it comes to identifying anomalous activity. To provide analysis as a network grows, vendors sometimes add more data sources at no cost and then charge clients based on the volume of analyzed data.

Vendors continually increase the volume of historical data captured and stored and add targeted solutions behind that aggregate data, and still deliver a product that continuously generates false positives.

How MixMode Develops a Baseline

MixMode harnesses the power of third-wave AI to evaluate, sort, and understand standard network behavior over a specific period. The platform analyzes network elements over a variety of timescales, the same way a human analyst would, but with the benefit of massive computational power. The result is an efficient, accurate, continual analysis of network behavior.

The MixMode platform seeks out patterns of interaction over many such time periods and then contrasts the pattern over the next short interval of five minutes with what was seen previously. If there is a deviation in the patterns, the platform classifies the security risk. Low-risk deviations do not trigger the typical deluge of alerts commonly triggered by most security platforms, eliminating the burden of analyzing (or worse, ignoring) scores of false positives. Once the platform establishes the baseline, it adjusts automatically on an ongoing basis, eliminating the need for ongoing tuning as is common with other security solutions.

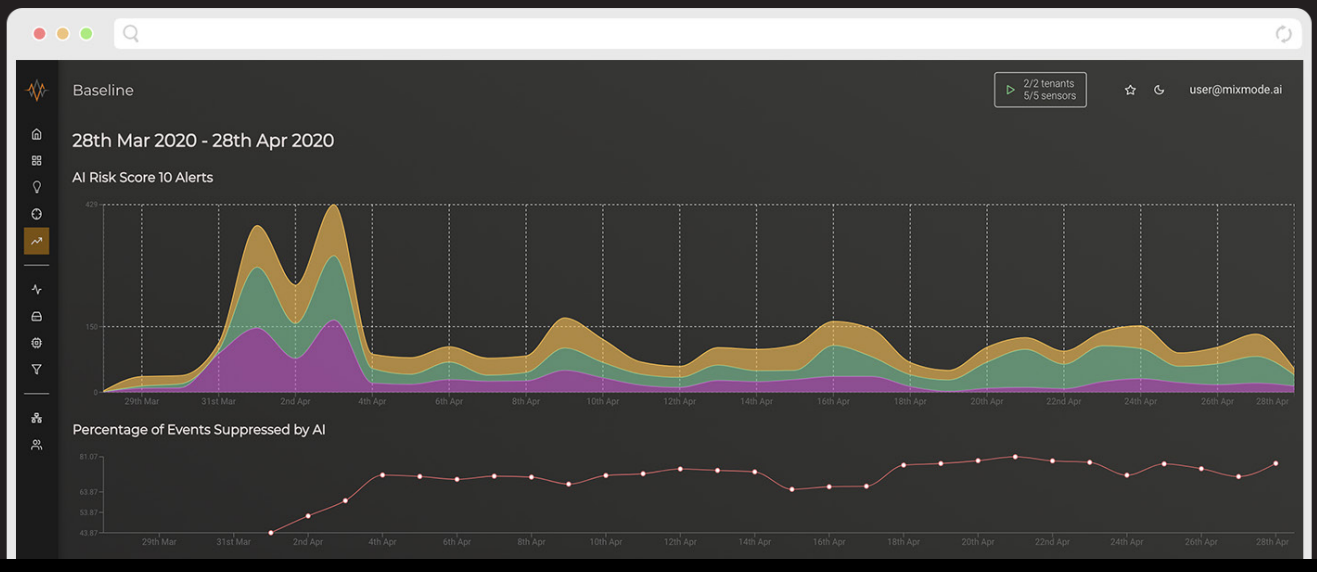
Coulehan says MixMode establishes a live, production-baseline within a week, a much quicker turnaround than the typical months required by most SIEM vendors. “Within a week, an organization can be surfacing and correlating traditional events and indicators and, most importantly, anomalous behavior that deviates from normal network traffic in a correlative way without human intervention or manual tuning,” he says.

Third-Wave AI Is Changing the Cybersecurity Landscape

Modern security threats require modern solutions. Second-wave AI has enhanced cybersecurity as networking has become a given across virtually all enterprises. However, second-wave AI functionality is not sufficient against sophisticated bad actors who have a solid understanding of how these platforms work.

MixMode utilizes third-wave AI to deliver a robust security solution that is changing the way enterprises handle Network Traffic Analysis across the board. By implementing MixMode, companies have an opportunity to add next-generation AI to their program, making their entire program more intelligent, efficient and productive. As Coulehan says, “It’s fascinating, it’s cost-effective, it’s resource-effective, and far less infrastructure intensive. It’s simply a better way of doing things.”

MixMode’s Powerful Patented AI allows users to create an evolving 7-day baseline of normal network behavior in minutes



MixMode™

www.mixmode.ai

+1 (858) 225-2352 | info@mixmode.ai |