Unveiling Hidden Threats with ML-Powered User and Entity Behavior Analytics (UEBA)

Avinash Gupta Desetty
Senior Splunk Engineer
Sony Corporation of America
gupta.splunker@gmail.com

ABSTRACT

The ever-growing cost of cybercrime has created the need for proactive solutions for organizations seeking to protect their digital assets. While traditional security systems struggle to detect anomalies buried within vast datasets, new solutions like User and Entity Behavior Analytics (UEBA) emerge as a game-changer. By leveraging the power of machine learning, UEBA analyzes diverse data sources like user logins, file accesses, event logs, business context, external threat intelligence, and network activity, to unveil hidden threats most traditional methods could miss. The ability to analyze multiple data sources enables UEBA solutions to effectively detect malicious insiders, compromised users, Advanced Persistent Threats (APTs), and zero-day attacks. By using various analytics techniques like supervised learning, unsupervised learning, and statistical modeling, UEBA solutions can detect subtle anomalies that deviate from established behavior baselines. Despite the many benefits, UEBA solutions still have limitations like data quality concerns, high implementation costs, and the need for model maintenance. Integration with System Information and Event Management (SIEM) systems helps mitigate some of these challenges to further enhance UEBA's capabilities and provide a unified platform for threat identification and response. This paper provides a detailed insight into the capabilities of UEBA, its three pillars, significance, and limitations.

Keywords: User and Entity Behavior Analytics (UEBA), Cybersecurity, Machine Learning, Threat Detection, Anomaly Detection, and Data Analytics.

Background

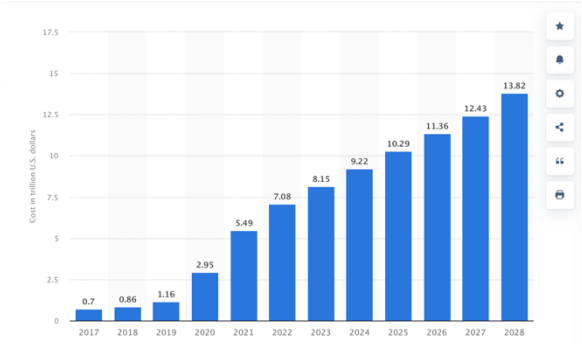

Figure 1 Estimate Cost of Cybercrime (2017 to 2028)

The cost of cybercrime increased by over 314% between 2017 and 2020, making it one of the major

costs for organizations that are not prepared [1]. Cybercriminals are leveraging technological advancements to execute even more sophisticated attacks. As threats become more sophisticated to detect, organizations need to counter them with even more sophisticated security tools that can analyze vast amounts of data and signals to detect threats in real time. Traditional security solutions often struggle to detect anomalies that lie hidden within the vast datasets of user and entity behavior. This is where User and Entity Behavior Analytics (UEBA) emerges as a game-changer.

UEBA leverages the power of machine learning (ML) to analyze vast amounts of data from diverse sources, including user logins, file accesses, network activity, and application usage. This approach was first coined by Gartner in 2015 as an advanced solution to solve the limitations of User Behavior Analytics (UBA). UEBA solutions can come in stand-alone tools or can be integrated into other security tools to make them more effective at detecting threats based on user behavior.

The need for UEBA has become increasingly critical in the face of several factors. Firstly, the rise of advanced persistent threats (APTs) creates the need for proactive detection methods [14]. APTs are often stealthy, carefully blending in with normal user activity, making them difficult to detect using traditional signature-based approaches. Secondly, the growing complexity of IT environments, with numerous users, devices, and applications, creates a vast attack surface for malicious actors to exploit. Manually sifting through the immense volume of data to identify subtle anomalies is simply not feasible.

This paper will explore how UEBA solutions leverage the power of AI and machine learning to detect hidden threats in an organization’s IT infrastructure. It will also discuss the limitations and challenges of UEBA solutions that users need to be aware of before integrating them into their security toolkit.

Literature Review

Threat detection is one of the fundamental pillars of any cybersecurity solution. In Barbara Filkins' comprehensive research (The Expanding Role of Data Analytics in Threat Detection, 2015), three major threat detection methodologies were discussed: Signature-Based (or Misuse) Detection, Anomaly-Based (or Behavior-Based) Threat Detection, and Continuous System Health Monitoring [2].

1. Signature-Based Detection

This method of threat detection relies on predefined patterns or signatures characteristic of known and documented attacks [3]. By analyzing network traffic through methods like deep packet inspection, potential signatures are extracted and cross-referenced against an already-established signature database to determine if they’re normal or not. The major strength of this method is its ability to limit false positives and its relatively lower processing requirements since it deals with a limited amount of data. However, it operates within a constrained scope since only known threats with existing signatures can be detected. This approach also creates the need for constant updates to the signature database to mitigate any new attacks that might be captured.

2. Anomaly-Based Detection

Timothy J. Shimeall and Jonathan M., and Spring's book, Introduction to Information Security, points out that Anomaly-Based (or Behavior-Based) works on the principle that malicious activities diverge

significantly from normal behavioral patterns [4]. Anomaly-based detection systems establish models representing the 'normal' behavior across networks, systems, users, and devices, subsequently flagging any deviations from these established norms. This method showcases the advantage of detecting threats without prior knowledge of their specifics. However, the major weakness of this approach is that it yields too many false positives. The complexity of system training in dynamic environments and the computational intensity involved in its execution is also another challenge.

3. Continuous System Health Monitoring

With continuous System Health Monitoring, an active vigilance strategy is used, which involves monitoring key system performance indicators to identify suspicious trends or changes. By actively observing network protocols, bandwidth usage, and unexpected traffic spikes, this method detects anomalies that show signs of malicious activities. This approach aligns with mapping unique malicious behaviors by devising system-wide measures and comprehending the significance of identified variations.

The UEBA Approach

User and Entity Behavior Analytics (UEBA) is a newer threat detection methodology, that integrates both anomaly-base detection and continuous system health monitoring [5]. UEBA solutions gather diverse data types including user roles, activity, location, and security alerts. This comprehensive data collection also includes both historical and current activities, considering elements like resource usage, session duration, connectivity, and group behavior for anomaly comparison. Crucially, it dynamically updates to reflect changes in user data, like changes in roles or modified permissions.

Unlike simply flagging all anomalies, UEBA systems assess the potential impact of behavior. For example, if an activity involves fewer sensitive resources, it receives a lower impact score. However, actions involving highly sensitive data, like personally identifiable information, have a higher impact score. This approach enables security teams to prioritize responses, focusing on higher-impact behaviors. Additionally, the UEBA system can automatically impose restrictions or heighten authentication requirements for users exhibiting anomalous behavior.

Detecting Hidden Threats with UEBA Using Machine Learning

Gartner defines UEBA solutions based on three pillars that are crucial in detecting hidden and sophisticated threats: use cases, data sources, and analytics [6].

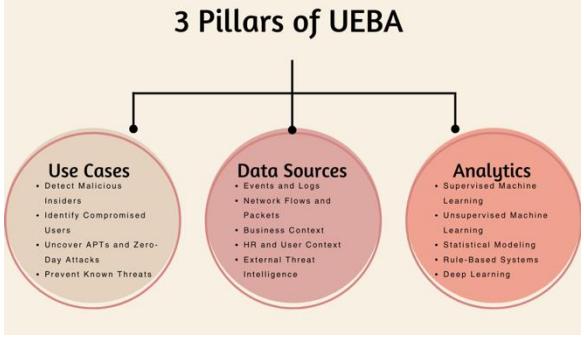

Figure 2 The Three Pillars of UEBA

Use Cases

UEBA solutions apply to multiple use cases, including the following;

- **Detect Malicious Insiders:** UEBA serves as a vigilant watchdog capable of pinpointing aberrant behavior within user accounts [13]. It analyzes patterns and activities, to recognize signs of potential malicious intent. For instance, it flags instances of unauthorized data access or any attempts to compromise system integrity.
- **Identify Compromised Users:** By analyzing changes in login patterns, access times, and data interaction, UEBA acts as a detective in identifying compromised user accounts [13]. It's good at detecting irregularities that might indicate unauthorized access or hijacking of user credentials, enabling rapid response to potential security breaches.
- **Uncover APTs and Zero-Day Attacks:** One of its most crucial roles lies in unearthing Advanced Persistent Threats (APTs) and zero-day attacks. UEBA's strength lies in its ability to detect subtle anomalies that deviate from established behavior baselines. Even in the face of unfamiliar attack techniques, it can identify these sophisticated, previously unknown threats.
- **Prevent Known Threats:** Leveraging machine learning models trained on historical data and threat intelligence feeds, UEBA operates proactively to thwart known threats. Whether it's detecting malware, thwarting phishing attempts, or mitigating brute-force attacks, UEBA's predictive capabilities are instrumental in fortifying defenses against these established threats.

Data Sources

UEBA solutions leverage a diverse array of data sources, each offering unique insights crucial for uncovering sophisticated threats:

- **Events and Logs:** These solutions delve into event and log data generated across various systems within an organization. This encompasses a wide range of activities, from user logins and file access to network interactions and security incidents. By analyzing this rich tapestry of data, UEBA unveils individual and collective behavior patterns, offering a holistic view of organizational activities.
- **Network Flows and Packets:** By scrutinizing network traffic data, UEBA identifies irregularities or anomalies within communication patterns. This scrutiny helps pinpoint potential threats like data exfiltration or botnet command-and-control communications. Analyzing network flows and packet data enables UEBA to detect subtle deviations that might indicate malicious activities lurking within the network.
- **Business Context:** Integration with business context data—such as financial transactions, customer information, and employee roles—enables UEBA to delve deeper into behavior analysis. This amalgamation allows for the identification of anomalies that might be inconspicuous in isolation. By contextualizing behavior within the business framework, UEBA gains a more comprehensive understanding of potential threats.
- **HR and User Context:** UEBA explores HR data and user profiles to glean insights into access privileges, roles, and historical behavior patterns. This contextual analysis provides a backdrop against which current activities can be evaluated. Understanding user behavior in relation to their roles and past actions aids in flagging suspicious deviations that might signal a security risk.
- **External Threat Intelligence:** Leveraging external threat intelligence feeds equips UEBA solutions with up-to-date knowledge about emerging threats, vulnerabilities, and evolving attack techniques. This ongoing influx of threat intelligence enables UEBA to adapt its detection models, focusing on the latest and most relevant threats to bolster the organization's security posture.

Analytics

UEBA solutions use several technical approaches to process and analyze vast amounts to detect threats. The following analytics techniques are used by UEBA solution;

- **Supervised Machine Learning**: UEBA solutions use supervised machine learning algorithms trained on labeled datasets [11]. These algorithms learn from known patterns to identify specific threats, such as account takeovers or data exfiltration. By recognizing patterns with signs of known threats, supervised learning enhances the precision of threat identification.

- **Unsupervised Machine Learning:** UEBA also uses unsupervised machine learning to detect clusters of anomalous behavior and outliers that might escape supervised methods [11]. These techniques excel in uncovering deviations that lack predefined labels, allowing for the identification of novel threats or irregular behavior patterns.
- **Statistical Modeling:** UEBA establishes baselines for normal behavior through statistical modeling [12]. By analyzing data distributions and establishing expected ranges, statistical models highlight deviations that fall outside these norms. This method effectively identifies potential threats by pinpointing activities significantly deviating from the established patterns.
- **Rule-Based Systems:** UEBA leverages pre-defined rules and thresholds to swiftly detect specific types of suspicious activities [13]. These rules provide an efficient means of identifying known threats. By setting predetermined criteria, rule-based systems swiftly flag activities that match predefined threat indicators.
- **Deep Learning:** Integration of deep learning algorithms empowers UEBA to extract intricate patterns and relationships from vast datasets. By delving into complex data representations, deep learning augments UEBA's potential to develop more sophisticated and powerful threat detection solutions.

Integration of UEBA Into SIEM Systems

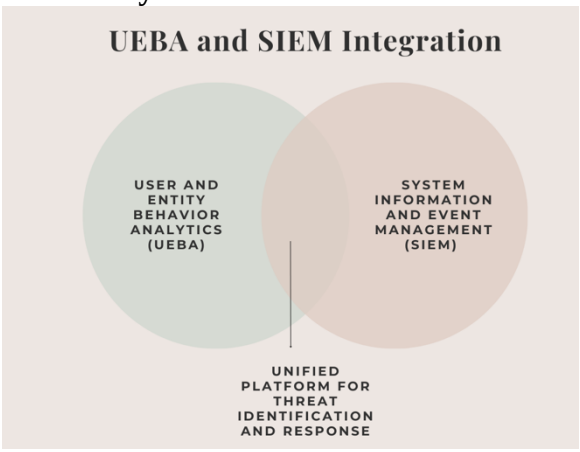

Figure 3 UEBA and SIEM Integration

The integration of UEBA capabilities into Security Information and Event Management (SIEM) was a pivotal advancement in cybersecurity strategies [7]. SIEM, known for its data aggregation and rule-based threat detection, collects and categorizes data from various network devices, using predefined rules to identify known threats. On the other hand, UEBA adds a layer of context and sophistication to SIEM by enhancing it with self-learning threat detection capabilities [7]. UEBA complements SIEM systems by offering a deeper understanding of behavior patterns, focusing on anomalies based on user and entity signatures rather than solely relying on rule-based detection.

The integration of SIEM's rule-based detection and UEBA's adaptive and contextual analysis forms a robust defense mechanism against evolving threats. Popular SIEM vendors, including popular ones like Splunk and Microsoft Sentinel, acknowledge the synergy between the two systems and offer integrated solutions that leverage both SIEM's established capabilities and UEBA's innovative approach [8]. This integration not only streamlines the process but also provides a unified platform that correlates related events, presents a comprehensive threat chain, and empowers analysts with a holistic view of potential security risks.

Significance of The UEBA Approach

UEBA improves in the following ways

Enhanced Detection Capabilities

The UEBA approach significantly enhances detection capabilities against sophisticated cyber-attacks, especially those originating from within an organization [9]. By leveraging automation and Machine Learning, UEBA conducts a comprehensive contextual analysis of vast amounts of raw data. This allows for the identification of complex behavioral patterns that might escape human analysis or traditional threat detection approaches. As a result, a broader spectrum of behaviors can be modeled and compared with peer groups, enabling the detection and mitigation of a wider range of cyber threats. For instance, UEBA proves instrumental in addressing privileged account exploitation, privilege escalation, and data exfiltration.

Quick Threat Mitigation

The speed and efficiency of UEBA in analysis and alerting are critical in an attack scenario, including insider threats [10]. UEBA's swift identification and alerting capabilities play a crucial role in minimizing Mean Time To Repair (MTTR). Rapid response is essential to limit an attacker's access to the network, thereby reducing the potential exfiltration of sensitive data.

Non-Disruptive Technical Implementation

From a technical perspective, UEBA operates seamlessly without disrupting legitimate network traffic or normal user behavior. This non-interference characteristic ensures that business operations remain uninterrupted. By design, UEBA's technical implementation functions effectively within the network infrastructure without causing disruptions, allowing for continuous operation and analysis without hindering day-to-day activities.

Optimized Human Resource Utilization

UEBA's analysis optimizes the utilization of IT resources within an organization. The automated detection mechanisms help alleviate the scarcity of experienced cybersecurity analysts by allowing skilled profiles to focus on critical analytical tasks [9]. This enables cybersecurity teams to concentrate more effectively on threat detection and response activities. Furthermore, UEBA's precise analysis of behavior significantly reduces false positives, distinguishing between different user roles and behaviors. The reduction in false alarms restores operational availability to security analyst teams, enabling them to tackle more complex and higher-value tasks

Limitations of the UEBA Approach

- **Data Quality and Completeness**: UEBA's efficacy heavily relies on the quality and completeness of the data it analyzes. Inaccurate or incomplete data inputs can result in false positives or overlook genuine threats. Ensuring robust data collection and management processes becomes crucial for providing UEBA with accurate insights.
- **Model Tuning and Maintenance:** Continuous tuning and maintenance of UEBA models are crucial to keeping pace with evolving threats and changes in user behavior. This demands expertise in machine learning and data science, skills that might not be universally available within most organizations.
- **Alert Fatigue:** In some instances, UEBA solutions may generate a multitude of alerts, potentially overwhelming security teams and leading to alert fatigue. Establishing effective alert management processes is crucial to prioritize and investigate high-risk alerts efficiently. Implementing streamlined processes helps prevent the dilution of critical alerts within the noise of lower-priority notifications.
- **Limited Scope of Detection:** While effective in detecting hidden anomalies, UEBA is not a one-size-fits-all solution. It might not effectively detect certain attack types, like zero-day exploits or sophisticated social engineering tactics. That’s why adopting a comprehensive security strategy that combines UEBA with other complementary security such as SIEM tools and IDS solutions becomes necessary.
- **Privacy Concerns:** The extensive data collection and analysis inherent in UEBA raise pertinent privacy concerns. UEBA's reliance on analyzing substantial user data requires organizations to establish clear data privacy policies and adhere to relevant regulations. Ultimately, balancing the need for enhanced security with user privacy protection remains a critical consideration during UEBA implementation.

Conclusion

UEBA is a transformative solution, especially for increasingly sophisticated cyber threats. By leveraging the power of machine learning and integrating diverse data sources, UEBA empowers organizations to detect hidden threats that may not be noticed by traditional detection methods. Its comprehensive and contextual analysis of user and entity behavior provides a deeper understanding of potential security risks, enabling proactive and effective threat mitigation. While limitations exist in terms of data quality, model maintenance, and potential for alert fatigue, UEBA's overall benefits outweigh these challenges, making it a valuable tool for organizations seeking to strengthen their cybersecurity posture. The integration of UEBA with established solutions like SIEM is one of the ways to minimize its drawbacks. Integrating SIEM and UEBA offers a comprehensive and unified platform for threat identification and response.

Overall, investing in UEBA solutions and ensuring their effective implementation will be critical for organizations to navigate the ever-dynamic cybersecurity landscape and safeguard their digital assets in the future.

References

[1] Statista, "Estimated cost of cybercrime worldwide 2017-2028.' Available online: https://www.statista.com/forecasts/1280009/cost-cybercrime-worldwide
[2] Wade W., Barbara F., "The Expanding Role of Data Analytics in Threat Detection," October 2015.
[3] Michael R., "What is the difference between signature-based and behavior-based intrusion detection systems?" December 2020. Available online: https://accedian.com/blog/what-is-the-difference-between-signature-based-and-behavior-based-ids/
[4] Timothy J., Shimeall, Jonathan M., Spring, "Introduction to Information Security," 2014. Available online: https://www.sciencedirect.com/book/9781597499699/introduction-to-information-security
[5] IBM, "What is UEBA (user and entity behavior analytics)?" Available Online: https://www.ibm.com/topics/ueba
[6] Gartner, "Market Guide for User and Entity Behavior Analytics," May 2019. Available online: https://www.gartner.com/en/documents/3917096
[7] Jason C., Jay B., "UEBA: Canary in a Coal Mine," April 2017. Available online: https://securityintelligence.com/ueba-canary-in-a-coal-mine/
[8] Splunk, "4 Reasons to Add UBA to Your SIEM." Available online: https://www.splunk.com/en_us/form/4-reasons-to-add-uba-to-your-siem.html
[9] GateWatcher, " Benefits of a UEBA Approach." Available online: https://www.gatewatcher.com/en/lab/benefits-of-a-ueba-approach/
[10] Aujas, "How to Mitigate Insider Threats with SIEM & UEBA," July 2020. Available online: https://blog.aujas.com/how-to-mitigate-insider-threats-with-siem-ueba
[11] Oskar C., | Daniel N., "User and Entity Behavior Anomaly Detection using Network Traffic," 2017. Available online: https://www.diva-portal.org/smash/get/diva2:1113229/FULLTEXT02
[12] Derek L., "Applying data science to user and entity behavior analytics," 2016. Available online: https://dataanalytics.report/Resources/Whitepapers/a93a20c4-fc03-4692-9247-d662092726ed_wd2.PDF
[13] Exabeam, "What Is UEBA (User and Entity Behavior Analytics)?" Available online: https://www.exabeam.com/explainers/ueba/what-ueba-stands-for-and-a-5-minute-ueba-primer/
[14] Linan H., Quanyan Z., "A dynamic games approach to proactive defense strategies against Advanced Persistent Threats in cyber-physical systems," 2020. Available online: https://www.sciencedirect.com/science/article/abs/pii/S0167404819302020?via%3Dihub