ManageEngine
Log360

Understanding User and Entity Behavior Analytics (UEBA):

How machine learning can help secure your business

www.manageengine.com/log360

TABLE OF CONTENTS

The continuous sophistication of attacks

The world's first cyberattack took place in November 1988 when Robert Tappan Morris wrote the first ever distributed denial of service (DDoS) program. In the last 30 years, cyberattacks on businesses have become much more frequent, and sophisticated.

The adoption of digital transformation strategies and the Internet of Things (IoT) has led to an exponential rise in the number of vulnerable endpoints that can be exploited - there are millions of connected devices on the internet today. The repercussions are evident in some of the cybercrime statistics. There were 1244 data breaches and 447 million exposed records in 2018 alone.[1] And a University of Maryland study concluded that a hacking attempt occurs once every 39 seconds.[2] Juniper Research predicted that cybercrime will cost businesses over $2 trillion by 2019.[3]

Moreover, cyberattackers have become more creative in their attack methods using tactics such as smart phishing, fileless malware, and zero-day exploits. Smart phishing involves the hacker doing substantial background research before targeting a victim. Fileless malware attacks occur when hackers, instead of installing malware on host machines, use built-in system tools such as PowerShell and Windows Management Instrumentation (WMI) to gain privileged access and move laterally. It is more difficult to detect such attacks as they go unnoticed amongst the routine activities that administrators carry out using these same tools. A zero-day exploit happens when a hack occurs on the same day that a vulnerability is exposed. Due to these sophisticated methods, the mean time to identify (MTTI) an attack was as high as 197 days, and the mean time to contain (MTTC) was 69 days in 2018.[4]

Thus every organization needs effective security solutions to safeguard itself against threats. Early detection and resolution can save organizations a huge amount of money. Implementing a Security Information and Event Management (SIEM) solution that analyzes the network's activities and helps to detect attacks, and a User and Entity Behavior Analysis (UEBA) tool that uses machine learning (ML) to detect users' and entities' behavior anomalies can act as a multi-layered defense strategy.

SIEM and UEBA

SIEM solutions enable organizations to collect and store logs in a central location. They also leverage different traffic flow protocols to keep track of other network activities. This makes it extremely convenient for IT administrators to set thresholds and conditions for real-time alerting in case of security incidents. SIEM solutions also enable IT administrators to correlate a series of events together to identify a threat that otherwise would have been missed. These solutions rely on known patterns or "signatures" to identify a threat vector. These signatures need to be fed into the system by administrators either through conditions or correlation rules, or through the automatic retrieval of data from threat intelligence databases such as STIX and TAXII. However, even after all of this, it is not possible to stop all attacks. Kevin Mitnick, arguably the world's most famous hacker, says, "You can never protect yourself 100 percent. What you do is protect yourself as much as possible and mitigate risk to an acceptable degree. You can never remove all risk."[5] SIEM solutions also help companies to take action even if an attack does take place. Effective techniques such as Elasticsearch can be used to perform forensic analysis and get to the bottom of why an event of interest occurred. The IT administrators can then take measures to ensure that the problem is resolved and an attack does not take place due to the same root cause again.

UEBA uses the power of ML algorithms to detect anomalies in the behavior of both users and devices on a network. These algorithms use statistical and probability models to establish a normal profile for each user or entity in an environment. Each action performed by a user or entity is compared to their profile generated by one of these models, using historical data. If an event doesn't fit in the list of what's expected, it's immediately classified as an anomaly and this information is given to the administrator who can then take appropriate action. ML-based defense systems learn on their own; their ability to defend against cyber threats increases as they gain experience. A UEBA system needs at least one day of historical data to start working, and at least two weeks of historical data to start working effectively.

There are several benefits of using UEBA:

1 Since the actions of each user and entity is compared to their corresponding baseline or "average", the number of false-positives and false-negatives will be reduced when compared to the rule-based alerting mechanisms.

2 While SIEM solutions treat security mishaps as isolated incidents and give alerts, UEBA solutions look at security holistically and calculate risk scores for each user.

3 It can offer better protection against zero-day exploits for which there are no known "signatures" yet.

4 An attacker could lurk in the network for a long time, pivoting from machine to machine, and gradually increasing their privileges, to go unnoticed. A UEBA solution can detect such long-term malicious lateral movements more effectively than SIEM solutions. The concept of risk scoring (which is covered later) will be of immense value here.

5 There is no reliance on IT administrators to develop thresholds or correlation rules to identify threats.

SIEM and UEBA are converging

SIEM and UEBA are converging as the days go by. Numerous SIEM vendors are developing UEBA capabilities, and numerous standalone UEBA vendors are starting to integrate their tool with SIEM solutions. While SIEM solutions can help detect known attacks and fix the problem as soon as possible with an integrated incident management response and workflows, UEBA solutions can help us detect the more sophisticated attacks. Together, they help security administrators effectively handle different threat scenarios.

How to choose security solutions?

The most important criteria for choosing a security solution should be based on specific use cases and pain points that an organization wants to address. An integrated SIEM and UEBA solution may help address several pain points. However, each company needs to assess their requirements and check if the solutions can be tuned to meet their security requirements.

Of course cost and return on investment is another criteria to keep in mind. According to the Gordon-Loeb model, a company should spend no more than 37% of the expected losses due to a cyberattack, on cybersecurity solutions.

How does UEBA work under the hood

Unsupervised ML is arguably the best way to detect anomalies. The UEBA system undergoes a "training" period during which it learns the baseline behavior of every user and entity. In case an anomaly is detected, the IT administrator gets the information on their dashboard. There is also a variant called supervised ML in which the UEBA system is fed the list of known good and bad behaviors. The tool builds upon these inputs and then detects different types of bad behaviors when they occur.

There are two main techniques or statistical models to decipher anamolous behaviors in a network: **1) Robust principal component analysis,** and **2) Markov chains.**

Robust principal component analysis (RPCA)

Principal component analysis (PCA) is a statistical method that finds the direction of the line of best fit for a set of observed data points (See Figure 1).

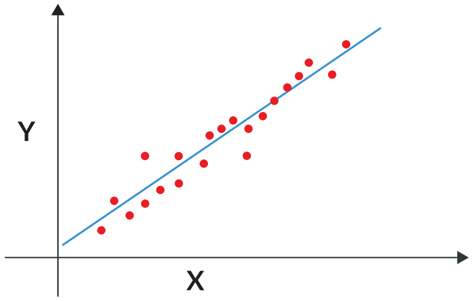


Figure 1: RPCA finds the direction of the "line of best fit" for a set of data points

Markov chains

A Markov chain is a sequence of stochastic events where the probability of the next event in a chain is depends only on the state of the current event. Markov chains typically ignore time as a factor. This model is often used to solve business problems in marketing and finance. For example, given a particular initial webpage a user navigates to (State 1), what are the probable successive webpages they can go to (States 2, 3, and so on)? And what is the probability of users attaining each of these states?

The probabilities of successive states are calculated to determine how risky a particular behavior is. Each action performed by a user or entity is compared to a list of probable actions. If an event is not found in the list of probable events, the UEBA system would see this action as an anomaly and raise an alert.

For example, given that a user has already failed to logon twice, what is the probability that this user will logon correctly on the third attempt? And what is the probability that this user will then access a database server and download important customer information onto a USB drive? Based on past behavior, the UEBA system will calculate the probability for each subsequent state, and give a risk score. See Figure 2.

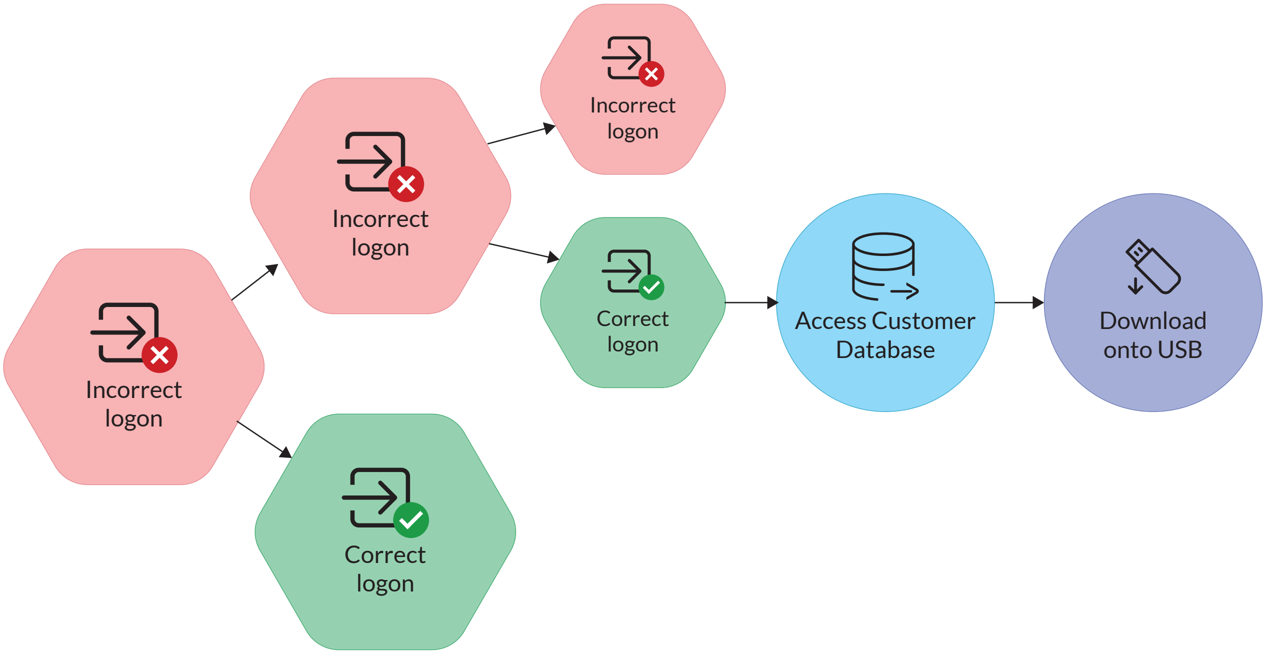


Figure 2: How ML using markov chains work

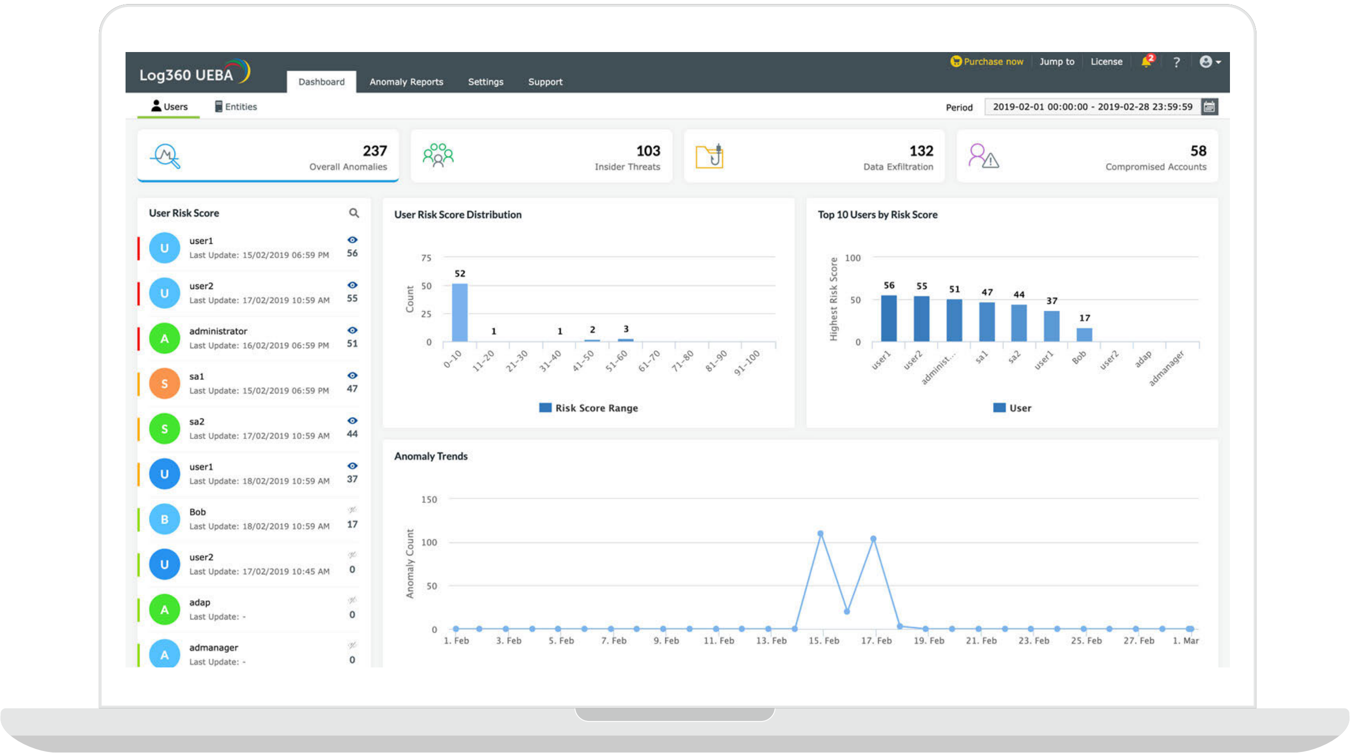
How ManageEngine Log360 uses ML techniques

ManageEngine's Log360 is an integrated SIEM and UEBA solution that can solve all of a business's IT security challenges. Log360's UEBA add-on can identify anomalous user and entity behaviors based on abnormalities in time, count, and patterns. The below table shows some examples of each type of anomaly, and the algorithm used for detection.

Type of anomaly	User anomaly	Entity anomaly	Algorithm used
Irregular time	An employee who generally logs on between 9am and 10am suddenly logs on at 5am. This will constitute a time anomaly.	A file is modified on a particular host at a time that is out of the expected range for that machine. Eg. unusual file modifications take place between 8 and 8:15 am on a machine, when the expected time range is between 4 and 4:15 pm.	RPCA
Abnormal patterns	A user logs on to a host that he has never logged on to before.	A server is accessed from a remote location for the very first time.	Markov chains
Irregular count	A user has executed over 20 DML queries on a SQL server while the baseline is usually three. This behavior will trigger a count anomaly.	A particular router has over 50 configuration changes when the expected number is only 13.	RPCA

Working with risk scores in Log360

An organization may face several internal and external threats. Log360 features a score-based risk assessment to help IT administrators prioritize threats and determine which events actually merit investigation. See Figure 3.



Period 2019-02-01 00:00:00 - 2019-02-28 23:59:59

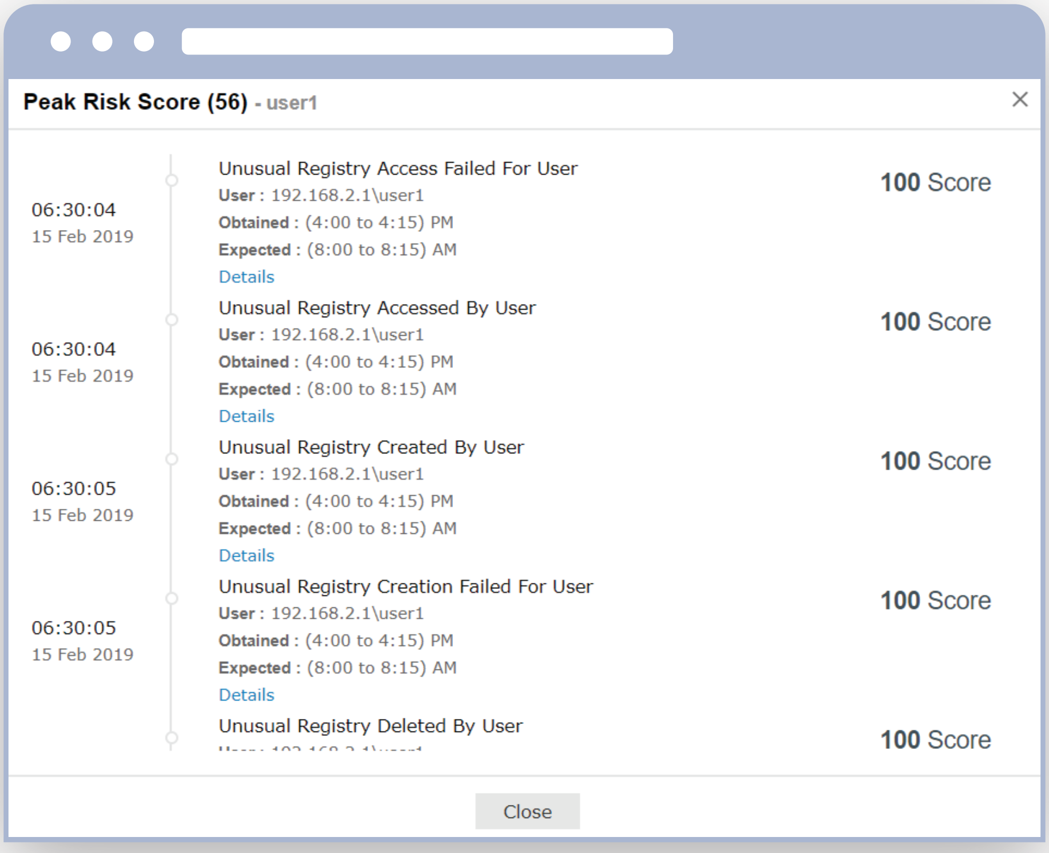
User Risk Score

U	user1 Last Update: 15/02/2019 06:59 PM	56
U	user2 Last Update: 17/02/2019 10:59 AM	55
A	administrator Last Update: 16/02/2019 06:59 PM	51
S	sa1 Last Update: 15/02/2019 06:59 PM	47
S	sa2 Last Update: 17/02/2019 10:59 AM	44
U	user1 Last Update: 18/02/2019 10:59 AM	37
B	Bob Last Update: 18/02/2019 10:59 AM	17
U	user2 Last Update: 17/02/2019 10:45 AM	0
A	adap Last Update: -	0
A	admanager Last Update: -	0

Figure 3: Log360 shows a risk score for each user and entity

The basis of this assessment is that all anomalies are assigned specific values, and a risk score is generated for each user and entity based on how dangerous their behavior is. The more unusual actions a user performs, the higher their risk score will be. Furthermore, risk scores can be connected to specific attack scenarios such as insider attacks, data exfiltration and account compromise. For instance, a user who has 10 failed logon attempts, followed by a successful logon will be classified as "high risk" for account compromise, and "not so high risk" for insider attack. This is because the chances of an insider failing to logon 10 consecutive times is low.

Users and entities with an increasing risk score can be added to a watch list to keep track of all their activities. Figure 4 shows an example of how an effective UEBA solution keeps track of all activities performed by watchlisted users.



Peak Risk Score (56) - user1

06:30:04
15 Feb 2019

Unusual Registry Access Failed For User
User : 192.168.2.1\user1
Obtained : (4:00 to 4:15) PM
Expected : (8:00 to 8:15) AM
Details

100 Score

06:30:04
15 Feb 2019

Unusual Registry Accessed By User
User : 192.168.2.1\user1
Obtained : (4:00 to 4:15) PM
Expected : (8:00 to 8:15) AM
Details

100 Score

06:30:05
15 Feb 2019

Unusual Registry Created By User
User : 192.168.2.1\user1
Obtained : (4:00 to 4:15) PM
Expected : (8:00 to 8:15) AM
Details

100 Score

06:30:05
15 Feb 2019

Unusual Registry Creation Failed For User
User : 192.168.2.1\user1
Obtained : (4:00 to 4:15) PM
Expected : (8:00 to 8:15) AM
Details

100 Score

Unusual Registry Deleted By User

100 Score

Figure 4: Log360 enables you to drill into each user action and associated risk score

Developing a risk appetite

Companies should gauge themselves on what level of risk they're comfortable taking. Do they want to address all the risks or do they only want to address risks identified as high? And what minimum risk score constitutes high risk? The answer to this question will vary from company to company.

Use cases in Log360's UEBA

As mentioned previously, the most important criteria for choosing a security solution should be based on specific use cases and pain points that an organization wants to address. ManageEngine Log360 UEBA add-on enables companies to address security lapses in three different areas: insider threats, account compromise and data exfiltration.

1

Insider threats

Both current and former employees need to be constantly monitored to ensure that the risk of insider threat is minimized. Abnormal login times, unusual file access, abnormally high number of file modifications, and high number of file downloads can all be indicators of an insider threat.

2

Account compromise

Apart from insider threats, an organization must also be wary of external hackers who gain access to the company network. Unusual number of logon failures, an abnormal logon from a remote location or host, or an abnormal denied connection on a firewall could be indicators of an account compromise.

3

Data exfiltration

According to Techopedia, data exfiltration is the unauthorized copying, transfer or retrieval of data from a computer or server. The risk score for data exfiltration should be higher when there there are multiple abnormal file reads by a user, an unusually high number of file downloads, or when an USB is plugged in an unusual time after an unusual file access.

Let's look at some real-world examples:

Compromised work station and data exfiltration attempt

Imagine a scenario in which an attacker gains access to an organization's network through a phishing email. This is a likely flow of events as they gain this unauthorized access, along with actions performed by Log360 (see Figure 5).

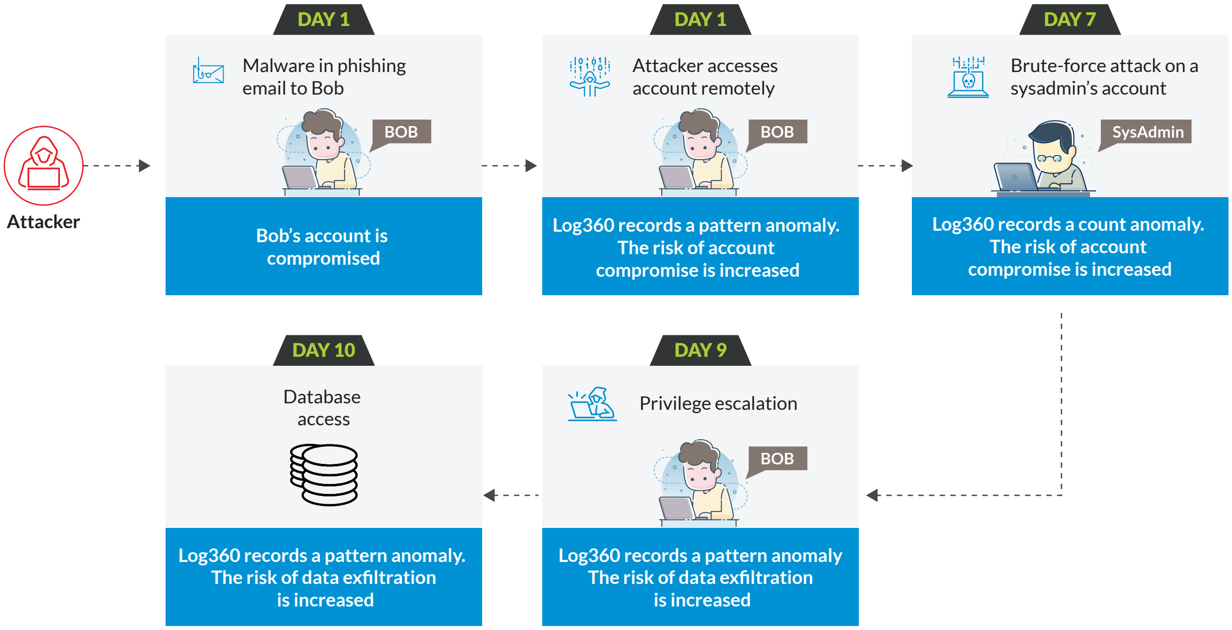

Attacker
DAY 1
Malware in phishing email to Bob
BOB
Bob's account is compromised
DAY 1
Attacker accesses account remotely
BOB
Log360 records a pattern anomaly. The risk of account compromise is increased
DAY 7
Brute-force attack on a sysadmin's account
SysAdmin
Log360 records a count anomaly. The risk of account compromise is increased
DAY 9
Privilege escalation
BOB
Log360 records a pattern anomaly The risk of data exfiltration is increased
DAY 10
Database access
Log360 records a pattern anomaly. The risk of data exfiltration is increased

Figure 5: An account compromise followed by a data exfiltration

1 On Day 1, the attacker sends a legitimate-looking phishing email to a front-end employee named Bob. The email contains a .doc file that Bob clicks on.

2 Without Bob's knowledge, malware starts being downloaded in the background. The attacker now has the knowledge of Bob's credentials.

3 The attacker connects to the network from a remote location using Bob's credentials. This will trigger a pattern anomaly and increase the risk score associated with an account compromise.

4 On Day 7, the hacker gets a hold of a system administrator's credentials using a brute-force attack. They have 10 failed logons followed by a successful logon on the 11th attempt, all within 30 minutes. This will trigger both a count anomaly and a time anomaly, and will in turn raise that administrator's risk score associated with an account compromise. The risk score associated with the specific machine that was used to carry out this attack will also be raised.

5 On Day 9, the attacker uses the sysadmin's credentials to escalate Bob's privileges. Bob is made a member of a privileged security group. This will again increase the "risk score" of a possible account compromise.

6 On Day 10, the attacker accesses an unusually high number of databases with confidential data using Bob's credentials. This will increase the risk score associated with data exfiltrations.

Insider threat and data exfiltration attempt

Imagine a scenario in which an engineer is told that their contract is not being renewed. The employee may become disgruntled and try to extract revenge on the company. This is a likely flow of events as they go about doing this, along with actions performed by Log360 (see Figure 6).

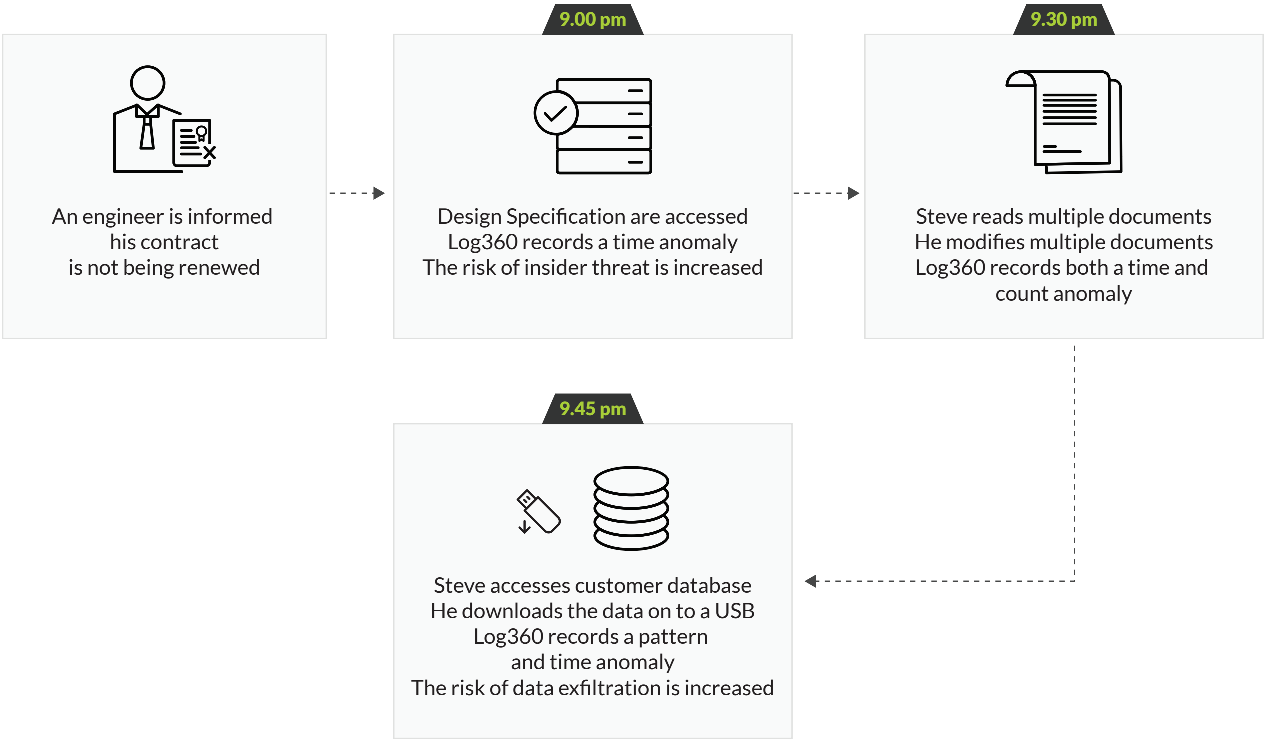

An engineer is informed
his contract
is not being renewed
9.00 pm
Design Specification are accessed
Log360 records a time anomaly
The risk of insider threat is increased
9.30 pm
Steve reads multiple documents
He modifies multiple documents
Log360 records both a time and
count anomaly
9.45 pm
Steve accesses customer database
He downloads the data on to a USB
Log360 records a pattern
and time anomaly
The risk of data exfiltration is increased

Figure 6: Insider threat followed by a data exfiltration

1 An engineer named Steve is told that his contract is not being renewed. The contractor decides to stay after his usual working hours and extract some revenge.

2 At 9pm, Steve accesses an important database that contains design specifications of a new product. Though he does have permissions to access this document, a time anomaly will be triggered due to the unusual access attempt time. The risk score associated with insider threats will also increase.

3 By 9:30 pm, Steve reads multiple design documents that he has access to and makes modifications to them. This will again raise his risk score due to the time and count anomalies.

4 By 9:45 pm, Steve gains access to the customer database, plugs in a USB flash drive and downloads the data. This will raise his risk score associated with a data exfiltration.

The future of UEBA in ManageEngine

Here are a few things to look forward to from ManageEngine when it comes to UEBA.

General AI:

Despite various advances in ML technologies, the IT industry hasn't developed machines that can learn from their varied experiences with zero supervision. General Artificial Intelligence remains a distant dream in the current day and age. However, there could well be a day when defensive machines can access the internet, learn about new attack patterns, and then incorporate these findings as they go about doing their job. This remains a vision of ManageEngine.

Reinforced or semi-supervised ML:

ManageEngine is also currently working on reinforced or semisupervised ML algorithms. This is a slight variant of unsupervised ML in which an IT administrator can can give feedback to the UEBA solution about its alerts. The tool will thus "learn on the fly" and its predictions will become more accurate.

Peer groups:

Clusters of users can be put into multiple groups based on their "average" attributes. A baseline can thus be calculated for each peer group instead of for each user. For example, employees from the marketing department can all be a part of one peer group, and employees from the finance department can all be a part of another peer group. Even if a particular employee from the marketing department has never previously logged on to the network at 8pm, it may still not be treated as an anomaly since it could be within expectations for the peer group.

Risk score customization:

In the near future, Log360 will allow users to customize the calculation of risk scores. Since every organization's nature of business is different, they may wish to give different weights to different types of anomalies. What constitutes a serious abnormality in one company may not be so serious in another.

About ManageEngine Log360

Log360 is a unified SIEM solution with integrated DLP and CASB capabilities that detects, prioritizes, investigates and responds to security threats. Vigil IQ, the solution's TDIR module, combines threat intelligence, ML-based anomaly detection and rule-based attack detection techniques to detect sophisticated attacks, and it offers an incident management console for effectively remediating detected threats. Log360 provides holistic security visibility across on-premises, cloud and hybrid networks with its intuitive and advanced security analytics and monitoring capabilities. For more information about Log360, visit manageengine.com/log-management/ and follow the LinkedIn page for regular updates.

About ManageEngine

As the IT management division of Zoho Corporation, ManageEngine prioritizes flexible solutions that work for all businesses, regardless of size or budget. ManageEngine crafts comprehensive IT management software with a focus on making your job easier. Our 90+ products and free tools cover everything your IT needs, at prices you can afford.

From network and device management to security and service desk software, we're bringing IT together for an integrated, overarching approach to optimize your IT.

$ Get Quote ⤓ Download

☎ Toll Free: +1 844 649 7766 DID: US : +1-408-352-9254

✉ log360-support@manageengine.com

🖥 www.manageengine.com/log360

Endnotes

1 *Annual number of data breaches and exposed records in the United States from 2005 to 2018 (in millions),*
Statista.,
https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed.

2 *Study: Hackers attack every 39 seconds,*
University of maryland.,
https://eng.umd.edu/news/story/study-hackers-attack-every-39-seconds.

3 *Cyber crime will cost businesses over $2 trillion in 2019,*
Juniper Research.,
https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion.

4 *2018 Cost of a data breach study: Global overview - July 2018,*
IBM.,
https://www.ibm.com/downloads/cas/861MNWN2.

5 *Kevin Mitnick quotes. Quoted in BrainyQuotes.,*
Kevin Mitnick.,
https://www.brainyquote.com/authors/kevin_mitnick.