User Behavior Analysis in the Cloud

Symantec

Security based on data science

User Behavior Analysis in the Cloud

Security based on data science

Contents

Symantec™

Introduction

CloudSOC™ uses the latest data science techniques, combining machine learning and advanced math, to provide fundamentally more intelligent and responsive security for the cloud. Our scientists are continually developing and tuning data science-driven engines and algorithms that take advantage of expansive processing and storage resources available in the cloud. This highly flexible scientific approach enables CloudSOC to keep up with the speed of change while identifying, analyzing, and controlling more user transactions with more cloud apps with more accuracy.

Cyber criminals specifically target cloud accounts as a means to access, infect, and breach organizations.

How do you detect cloud threat activity when it occurs outside your network infrastructure? How do you detect advanced threats or malicious user activity when no signatures exist to identify them? How do you analyze cloud traffic when cloud providers make it difficult to identify user transactions within network traffic?

Profoundly improved visibility and protection against compromised cloud accounts leveraging new data science techniques and elastic cloud resources

Traditional approaches may not provide the visibility or granular control that you need. However, new approaches open up new possibilities for solving security problems in the cloud. CloudSOC solutions are built upon innovative data science-driven engines that break through known constraints of traditional security solutions by leveraging the power of the cloud itself—leveraging cloud resources for more processing power and more flexible storage options.

Critical Intelligence Necessary to Protect Your Organization

A highly effective and accurate cloud security system must be able to identify key information and to evaluate the contextual significance of that information in order to turn it into useful intelligence. CloudSOC solutions are based on data science-driven engines to address intelligence on cloud transactions essential to effectively protect your organization.

CloudSOC uses data science to increase your knowledge and visibility over cloud activity:

- What is happening between your users and the cloud?
- What actions are your users taking in what cloud services?

CloudSOC uses data science to identify if you have a security issue:

- Is this activity a problem?
- Do I have an account compromised by hackers or malware?

Data Science for Better Visibility, Control, & Response to Threats

You need deep visibility into real-time traffic, not just what apps users are accessing, but also what exactly are they doing within that app. Getting to this level of granular and contextual knowledge is difficult. It requires a system with the ability to read the real meaning in volumes of traffic that uses obscure machine language identifiers to communicate with disparate systems. Additionally, this system must be adaptive, able to use a foundation of knowledge based on a continually learning system because these machine language identifiers can be changed without notice or documentation at any time by 3RD party cloud service development teams.

CloudSOC data scientists leverage the unique horsepower of cloud computing and machine learning to build a uniquely rich foundation of knowledge. Based on that foundation, they build contextual algorithms that can deliver a more detailed understanding of user behavior and cloud activity than possible with other traffic analysis systems. Then they leverage cloud processing power to execute these advanced algorithms.

Supervised Machine Learning

FOR WHEN YOU KNOW
WHAT YOU DON'T KNOW.

Supervised machine learning is a great way to analyze large quantities of source data and sort it into a foundation of knowledge that can be used by systems to make decisions and take actions. It enables a system to use a much larger set of source data, analyze it based on a larger set of characteristics, and process that big data to achieve more effective outcomes.

Unsupervised Machine Learning

FOR WHEN YOU KNOW
YOU DON'T KNOW
WHAT YOU DON'T KNOW.

Unsupervised machine learning lets the machines do freeform data discovery. It is a great way to discover source data necessary to guide learning systems to make smart decisions, when you don't specifically know what that source data should be.

StreamIQ is the advanced extraction technology that enables CloudSOC to understand transactions performed by users in a cloud app in more details than possible in most traffic analysis systems such as what is found in Next Generation Firewalls and Secure Web Gateways. This improved ability to identify the who, what, where, and when in traffic between your users and cloud accounts is critical to identifying and acting on potential threats to your organization.

Analyzing Cloud Activity with StreamIQ

In traffic analysis you need to track what activities are being performed by what users with what cloud apps in what context. You need details such as: What actions are being taken? Are these actions important or not? Are they associated with a specific file with specific attributes that would make them important? Are these actions associated with a cloud app you consider risky? Is this activity normal for this user?

New cloud apps are popping up all the time and existing cloud apps are continually changing their programming. Any system would find it extremely difficult to keep up with this constantly shifting environment. Identifying action indicators, object identifiers, and user information from machine readable text can be exceedingly difficult to identify. Traditional approaches can't keep up and as a result can track only a few gross identifiers and commonly break without warning when cloud services change their algorithms.

StreamIQ technology leverages both unsupervised and supervised machine learning and deep content inspection to extract granular cloud activity, which fuels CloudSOC's Protect, Detect and Investigate applications.

CloudSOC uses both unsupervised and supervised machine learning to create StreamIQ, the intelligence engine and algorithms that fuel CloudSOC traffic analysis, ThreatScores for Detect, Protect rules for visibility and policy control, and the high quality log data in Investigate for incident response. Our scientists start with a few significant characteristics known to be associated with important traffic attributes.

StreamIQ Intelligence Fuels Elastica CloudSOC Detect, Protect, and Investigate

Machine learning in StreamIQ drives more accurate and deeper real-time activity tracking for more cloud apps. Elastica solutions use the unique intelligence in StreamIQ to detect more threats, enforce protection with a more granular level of control, and investigate security incidents more effectively.

StreamIQ

- **Identifies more details on granular transactions in live traffic**
- **Analyzes traffic and identifies instructions custom to many apps—sanctioned and unsanctioned**
- **Automatically updates to accommodate cloud app code changes to stay accurate**
- **Powers more accurate risk analysis based on better activity intelligence**
- **Enables more granular policy controls**
- **Provides more useful data for incident response investigations**

StreamIQ Traffic Analysis

🟦	Application	filesharing.cloudapp
🟨	Action	Downloading multiple files as a ZIP file
🟧	Files	password.txt and id_rsa

```
POST https://12.dl.filesharing.cloudapp.com/documents/unshared?
session=KSGBYV8TQZX&t=zip&aqs=chrome..69i57.2678j0j1&sourceid=
chrome&ie=UTF-8 HTTP/1.1
Host : 12.dl.filesharing.cloudapp.com
cookie : PREF=ID=08DHMNG54O2X:U=2Q7SPLK15OTW
content-length : 126
user-agent : Mozilla/5.0 (Macintosh; Intel Mac OS X 10 _ 9 _ 2)
content-type : application/x-www-form-urlencoded;charset=UTF-8
accept : */*

token=9YDP70JR5ZCS&payload={[“file”:“passwords.txt”,
“parent”:“credentials”,”confirm”:false,”expires”:60},
“file”:“id _ rsa”,“parent”:“credentials”,”confirm”:false,
”expires”:60]}
```

Significance

The domain is cleared of portions (“12” and “dl”) that occlude the actual cloud app (“filesharing.cloudapp”).

The action (Downloading as a ZIP) is not explicitly stated and must be inferred from multiple portions of the URL. For this application, downloading as a ZIP indicates that there will be one or more files comprising the ZIP, and we should search for each of them.

The filenames are embedded in a hierarchy of data formats and are not near one another, increasing the difficulty of extracting them.

They use these as starting points for unsupervised machine learning that can identify significant instructions in machine code that would be very difficult or maybe impossible to find any other way. This foundational discovery of significant instructions is then fed into supervised machine learning systems that provide the content and contextual intelligence needed to turn this data into the foundation of knowledge in StreamIQ. Then powerful StreamIQ algorithms use this knowledge base to read traffic no other system can interpret. Because machines do this work fast, the CloudSOC system can keep up with a continually changing cloud landscape.

Essentially, StreamIQ figures out what the machine code in cloud traffic actually means thanks to this data science approach so it can deliver a uniquely granular level of traffic intelligence into the CloudSOC solutions.

Security that Recognizes Risky Activity

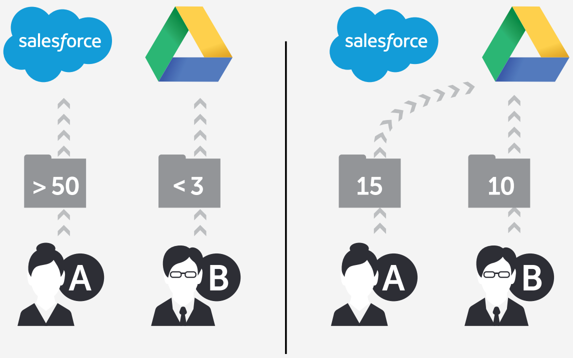
Once you know what is happening in cloud apps, you must be able to identify if that activity poses a risk. The key to activity-based security analysis lies in the ability to identify when activity represents abnormal user behavior likely indicating a threat. Cloud activity that follows typical user behavior patterns indicates everything is probably normal. Malicious activity, whether caused by a malware attack, a hijacked account, or a malicious insider, usually manifests abnormal activity that can be identified—for example, more frequent logins or uploads than normal for a particular user can indicate an account takeover. It may sound simple, but activating it effectively requires extensive foundational knowledge and smart adaptive tracking systems. Otherwise, you have a system that doesn’t identify abnormalities very well, or requires too much manual babysitting because it creates a lot of false positives.

User Behavior Analysis

Generic user behavior based security controls rely on manually set event thresholds and simple defined actions. This is not true user behavior analysis because these simplistic controls are not set based on individual user behavior. These are based on gross assumptions are relatively easy to set up but not very accurate, unless used judiciously and balanced by more nuanced user behavior analysis. An example of a useful generic behavior threshold control would be a rule to freeze access to an account if there were three failed user login attempts within a short period of time.

Another common generic threshold control is to trigger a response if a user uploads more than a certain number of files within a particular time period. But how do you decide what number constitutes larger than normal when some users hardly ever upload files and others upload lots of files? If this arbitrary threshold were too high it won't catch legitimately malicious activity, and if it were too low it will trigger lots of false positives creating extra work for IT and frustration for users.

For Example User A may typically batch upload 50 files every Friday to Salesforce, but never uploads files to Google Drive, except one day when they batch upload 15 files. User B may rarely upload files to Google Drive except one afternoon when they suddenly upload 10 files.



A generic user behavior threshold based on 20 uploads in 10 minutes would falsely flag User A behavior with Salesforce as potentially malicious, but not flag the Google Drive uploads and wouldn't register User B behavior as abnormal at all.

User Behavior Analysis Intelligence Feeds Elastica CloudSOC Detect and Protect

Machine learning enables highly granular personal user behavior profiles to more accurately identify risky activity in cloud apps. CloudSOC solutions use intelligence based on user behavior analysis and ThreatScores to detect threats, automatically enforce policies and provide better visibility into risky activity.

CloudSOC UBA Intelligence

- **More aware of abnormal activity due to more granular understanding of typical user behavior**
- **Minimizes false positives though individualized and contextualized user behavior modeling**
- **Faster response with automated ThreatScore calculations**

The CloudSOC system is designed to identify individual user behavioral patterns in context with app, time, objects, access method, etc. This user-specific, context-based method is much more accurate for identifying potentially malicious activity. However, a system that can provide a unique baseline for individual user behavioral patterns requires the ability to classify, analyze, and maintain a large volume of intelligence data. It requires a system able to adapt to changing patterns over time, and able to interpret the significance of deviations from normal and translate that deviation as usable, actionable information.

Individualized & Contextualized User Behavior Profiles

CloudSOC uses machine learning with expansive cloud processing and storage resources to power a self-training User Behavioral Analysis (UBA) engine. The UBA engine uses computational analysis algorithms to analyze transactional data from StreamIQ and ContentIQ. UBA algorithms develop a confidence curve for normal behavior customized to individual users in context with specific actions, apps and other attributes to create and maintain collections of highly accurate user behavior profiles.

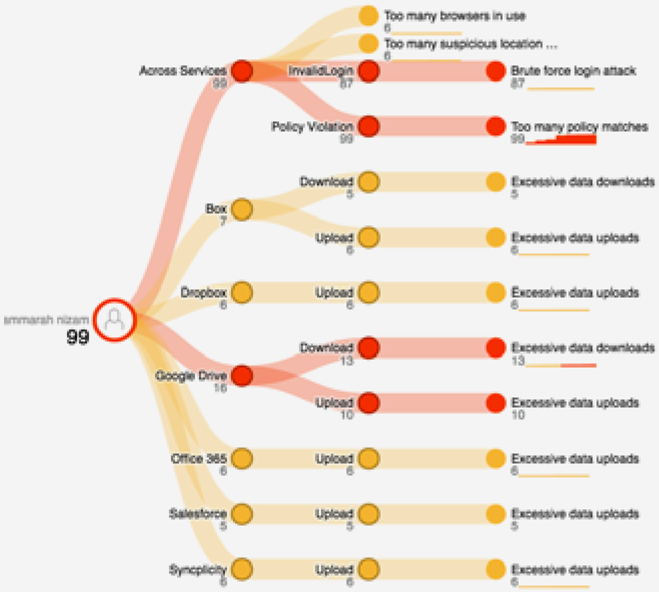
This foundation of knowledge baseline for normal activity opens up many more opportunities to accurately identify abnormal and potentially malicious activity without creating a deluge of false positives at the same time.

User A	B_1	C_1	D_1
User A	B_2	C_1	D_1
User A	B_2	C_2	D_1
User A	B_3	C_1	D_4
User A	B_5	C_2	D_1

Identifying Suspicious Activity with ThreatScore

Once a system can identify what is normal, it becomes possible to identify what is abnormal and therefore suspicious. If only it were so simple. How far from normal must behavior drift before it becomes abnormal? How do you evaluate increasing levels of risk as abnormal activity increases? How can you enable the solution to automatically respond with appropriate levels of security controls?

Our scientists tackled this problem with another layer of data science to identify and measure the severity of activity that deviates from normal. CloudSOC Detect uses computational analysis of user behavior to identify and score the severity of incidents representing risk. It then correlates this user behavior score with threshold-defined triggers and detection of suspicious sequences of events to calculate a dynamic, continually updated ThreatScore for each user and action.



CloudSOC Detect displays a dynamic map of user behavior events with granular event ThreatScores and color coding to identify levels of risk severity for each user.

The Information You Need for Incident Response

Security incidents will occur. That's the reality of today's cloud threat landscape and IT departments will at some point be scrambling to figure out what happened. This type of investigation can be challenging if not impossible with traditional perimeter security.

Typical Challenges Faced by Incident Response Investigations

Traditional data sources used to investigate security incidents offer up some big challenges, such as:

- Appliances with limited historical data due to storage resource constraints
- Log data that doesn't include enough granular information to answer important questions
- Vast quantities of redundant or irrelevant logs requiring lots of manual effort to glean useful information
- Logs full of data designed to be read by machines not humans making it difficult to interpret the data they contain
- Inability of on-premises appliances to monitor cloud usage or activities by mobile users

Well Designed Intelligence Engines and Cloud Resources to the Rescue

The limitations of traditional systems for incident response can be solved by leveraging the cloud, applying data science driven intelligence gathering, designing great algorithms that can interpret the data and a system that presents that data in an intuitive, easy to interpret format. This is what CloudSOC delivers.

Logs, your foundation of knowledge discovery for incident response, can only include the activity data that the original security system can read, so the quality of this data ultimately depends on the intelligence of your firewall, proxy, IPS, CASB or whatever system. If the underlying intelligence of a system can only read gross details in its traffic analysis, that's all you'll get from those logs. This is where the power of StreamIQ really shines.

StreamIQ picks up detailed activity data that other traffic analysis systems can't identify. Then it correlates activity details with multiple related attributes for contextual analysis and translates it from machine code to human language. This results in logs that are uniquely full of useful information and easy to understand. CloudSOC logs automatically consolidate multiple related less important actions under the one action of related contextual significance.

Office 365	Bob Jones sent an email to Alice Smith with the subject “Billings” using Exchange on April 12, 2016, 11:32 AM
Dropbox	ALERT bob@company.com attempted to Share book.xlsx using Linux and Firefox v43 on April 12, 2016 11:34 AM
Box	File “book.xlsx” has risk of PII and PCI violations from user bob@company.com
Google Drive	ALERT Bob Jones shared document “book.xlsx on April 12, 2016 11:45 AM
Office 365	Bob Jones user ThreatScore is now 97, changed for “Too many suspicious location changes” on April 12, 2016 11:59 AM

For example, StreamIQ data in logs tells you which user was involved instead of just presenting IP addresses, and it creates a record that this user logged into a particular account instead of creating multiple records separately tracking each step of the login process. You get logs that make it easy to track who was accessing what file in what app, what the attributes were of that file, what changes the user made to that file, and what permission settings were changed related to that file or account.

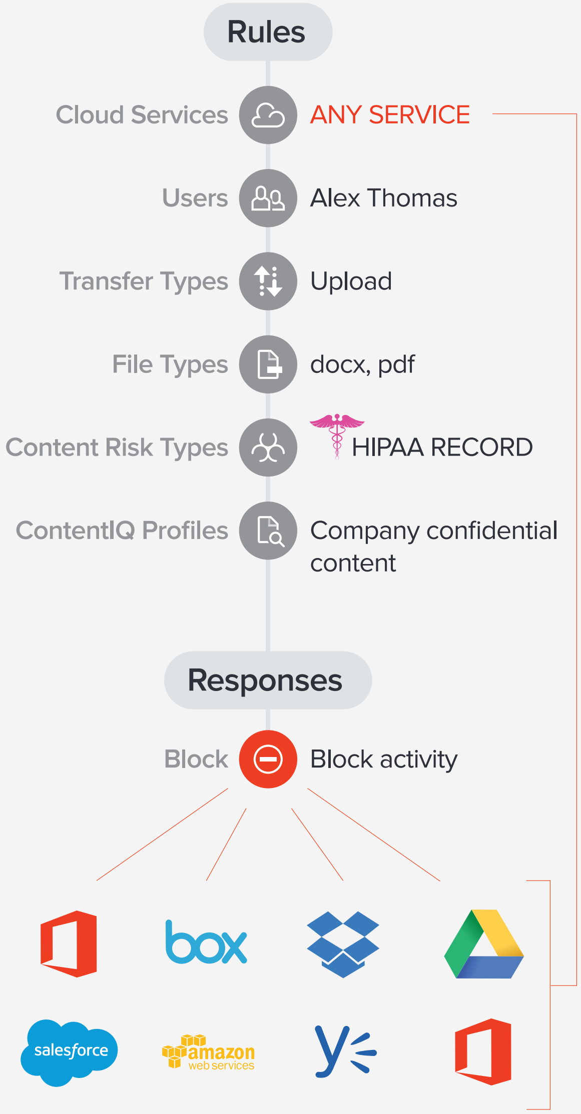
Pulling it All Together

The best threat intelligence in the world is useless if it can't find threats or interpret them in a timely manner. The first thing you'll notice when you get to the Investigate dashboard is a Query function. This is key, because wading through lots of irrelevant logs to find the ones you need is a waste of time. Investigate has a powerful but easy to use query where you can use a wide range of intuitive query terms combined with keywords to search by app, user, action, file, etc. Or you can skip query and use the rich set of data filtering options just beside the query feature.

The Investigate interface pivots based on the data returned from your query or filter settings. It automatically populates data visualizations and presents relevant logs full of drill down details thanks to all that intelligence work done by StreamIQ.

Data Science Based Policy Controls

Layers of data science driven systems from StreamIQ, to User Behavior Analysis, to ThreatScores make it possible for CloudSOC to provide visibility and control over cloud apps with an accuracy not possible with previous technologies.

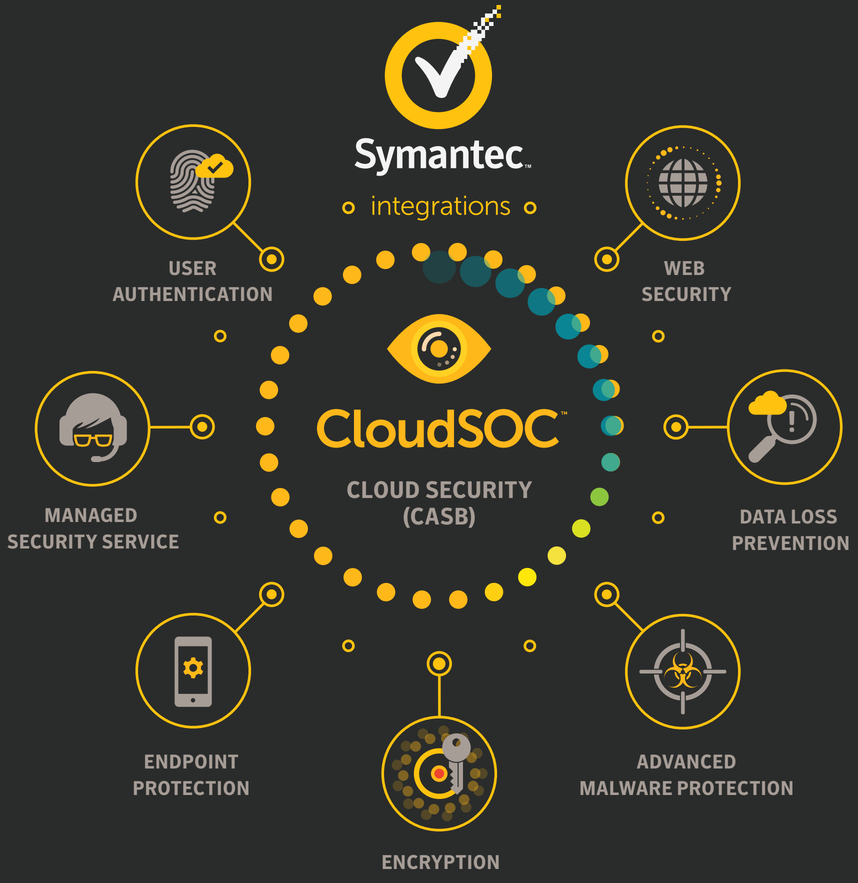


Conclusion

Cloud Security is built on a foundation of data science and cloud resources to deliver fundamentally better cloud security. This approach enables CloudSOC to move beyond many well known limitations of traditional security systems. Layers of machine learning, computational analysis, and intelligent algorithms go into building the highly accurate and adaptive StreamIQ and UBA engines at the core of CloudSOC. ThreatScores are calculated based on these engines to facilitate practical everyday security management, big data visualization, and automated controls.

Get better security with less complexity

Deploy a cloud security solution that integrates with your existing security infrastructure. A Symantec solution with CloudSOC provides greater security coverage, reduces operational complexity, and provides an optimal user experience.



For more info on Symantec CloudSOC CASB and its industry leading integrations with Symantec Enterprise Security Systems, visit **go.symantec.com/casb**

About CloudSOC

Data Science Powered™ Symantec CloudSOC platform empowers companies to confidently leverage cloud applications and services while staying safe, secure and compliant. A range of capabilities on the CloudSOC platform deliver the full life cycle of cloud application security, including auditing of shadow IT, real-time detection of intrusions and threats, protection against data loss and compliance violations, and investigation of historical account activity for post-incident analysis.

About Symantec

Symantec Corporation (NASDAQ: SYMC), the world's leading cyber security company, helps businesses, governments and people secure their most important data wherever it lives. Organizations across the world look to Symantec for strategic, integrated solutions to defend against sophisticated attacks across endpoints, cloud and infrastructure. Likewise, a global community of more than 50 million people and families rely on Symantec's Norton suite of products for protection at home and across all of their devices. Symantec operates one of the world's largest civilian cyber intelligence networks, allowing it to see and protect against the most advanced threats.

For additional information, please visit www.symantec.com or connect with us on Facebook, Twitter, and LinkedIn.

Symantec.

symantec.com +1 650-527-8000