Using HawkEye AP 6.0 for Detecting Insider Threats
Table of Contents

Why Discovery is as Important as Detection ................................................................. 3
The Insider Discovery Challenge ................................................................................ 3
Introducing HawkEye AP 6.0 Advanced Analytics .................................................... 4
What is a Data Model? ................................................................................................. 5
How We Find Insiders ................................................................................................. 5
   1. Configure the Conceptual .................................................................................... 5
   2. Collect the Data .................................................................................................. 6
   3. Execute the Data Flow Model ............................................................................ 6
   4. Visualize the Results ......................................................................................... 8
Conclusion .................................................................................................................. 9
Using HawkEye AP 6.0 for Detecting Insider Threats

Why Discovery is as Important as Detection

Today, significant amounts of IT Security expenditures go to detection including firewalls, anti-virus, SIEM, IDS/IPS, sandboxing and more. Yet time and time again threats get through the perimeter defenses without being detected. The latest Verizon Business Report suggests over 66% of breaches are discovered more than one month after initial compromise, and 69% of the time organizations are informed of the breach from an external source.

Organizations need to look at more than just the perimeter. They need to retain and look at all the historical activity going on in their environment. The vast majority of breaches are discoverable in the log data that machines generate. By saving and analyzing this log data organizations not only enable the ability to discover security breaches, but also obtain the information that is needed to investigate the nature of the breach, remediate, and possibly create new detection rules to stop similar breaches from occurring in the future.

There is a category of security breaches that can only be found by looking at historical events, and that is the category of Insider Threats. This includes Insider Threat, Insider Abuse, and some cases of Insider Fraud. It may or may not be intentional. This type of security breach defeats almost all types of detection by perimeter defenses since the threat is already inside the perimeter.

The Insider Discovery Challenge

Organizations can curtail behavior that they know is bad, either before it happens with OS and application access controls, or as it happens with detection tools such as DLP.

But what about new, unknown behaviors with unknown patterns and no signatures? There are behaviors that can be identified only in business terms, but when translated to computer and network behavior the specifics of the behavior changes as quickly as the IT landscape changes.

The challenge is creating a system with heuristics that operate at a high-level across historical activity from a wide variety of sources. These kinds of heuristics work best against deep historical information, going back six months or more. The heuristics need to work against as wide a variety of sources as possible. Ideally, the system should have a means of normalizing the data sources without compromising audit-ability so that your IT landscape can change and new systems will work in the same manner as other, similar systems. These kinds of heuristics leverage complex algorithms, so the system needs to be capable of advanced mathematical formulas and statistics. And the heuristic results need to be precise; it is most undesirable to have false positives with insider personnel.

The answer to this challenge: Yes, you do need big data for security.
Using HawkEye AP 6.0 for Detecting Insider Threats

Introducing HawkEye AP 6.0 Advanced Analytics

New in version 6.0 of HawkEye AP (Analytics Platform), Hexis has added advanced analytics data modeling. This is another layer of technology to help customers leverage the powerful Event Data Warehouse (EDW). With the advanced analytics layer customers can perform multi-step analytics that can be far more complex than any single query.

What is a Advanced Analytics Data Modeling?

A data model is a blueprint for manipulating data with multiple different steps. Each step performs an independent operation on the data, which may or may not be a single query against a database returning a single result set. The data may originate from an EDW query; however, it may also originate from a file, from an LDAP directory, from PostgreSQL or other database sources, or from user input. Multiple data sources can be annotated, augmented, merged, correlated or manipulated in arbitrary ways.

Models are created in a standard HTML web browser and are executed on the server of the Analyzer within a HawkEye AP deployment. Customers no longer need to export query results and import the results into other third-party tools, such as Microsoft Excel, to perform these complex manipulations.

Our insider threat discovery technology is merely one example of a model that can be created with the Advanced Analytics module.
How We Find Insiders

At a high level, the insider threat model begins with three different log sources, typically stored in EDW (e-mail behavior, Internet upload behavior, and login behavior) and one source, typically stored in an enterprise application (HR information). The relative importance of these four sources can be tuned for each specific customer implementation. The model then executes several statistical algorithms incorporating the weighted importance, and calculates a risk number for every individual in the organization. Finally, the results of the calculations are surfaced in an executive insider threat-risk dashboard. The details are as follows:

1. Configure the Conceptual Model

A base concept of this model is a hierarchy of behaviors that are each evaluated independently and then combined with a simple additive weighting algorithm. Currently the hierarchy is two levels deep, has four categories at the branch level and 17 categories at the leaf level.

Define Scoring for Behaviors

Throughout the context of this hierarchy weights are applied to define the relative importance of each behavior. The default weights are a strong recommendation from Hexis for initial deployments, however they can be adjusted for each organization to fine-tune the model to the specific needs of the organization. Weights are defined within each branch for the behaviors of all the leaves in that branch, and then again at the top to define the relative weight of each branch.

Define Anomaly Threshold (Z-Score)

Once the hierarchy is defined, the model executes, and each individual in the organization is scored for each day of their activity (referred to as the “IT-score” in reports). The scores of each individual are compared to the average score of everyone in the organization using the Z-Score (measurement of how far away an individual’s score is from the community average score). In addition, each individual user score is compared to that user’s own historical scores over a meaningful time period (30 day default) using the Z-score metric (i.e. a measure of how far an individual’s score is from that individual’s historical average). Metrics can be derived from the previous day’s average as well as the previous week, month, quarter and year’s averages.

The executive insider dashboard ranks the users based on the insider threat risk. By default, the historical trends of the top 3 risky accounts are displayed on the executive insider threat dashboard. This is a configurable threshold that can be changed to tune the model to an organization’s specific needs. For context, the historical trends of the daily average score and the daily score’s standard deviation detection threshold are included in the chart.
Using HawkEye AP 6.0 for Detecting Insider Threats

2. Collect the Data

Need Sources to Fill in the Conceptual Model
The ability to load any data source in the IT Infrastructure is a fundamental requirement for successful identification of insider threats. Organizations must be at liberty to use any and every obscure source that may reveal abnormal behavior. HawkEye AP supports hundreds of sources (see http://www.hexiscyber.com/sites/default/files/hexis-cyber-solutions-log-adapter-list.pdf) and is extraordinary at accommodating custom sources. Many organizations have internal home-built systems that are central to user activity. These logs can be critical for detecting abnormal behavior.

Need Original Raw Record for Audit and Normalized View for Analysis
If your organization discovers an individual who may be a threat, you want to be very sure the information you have is accurate before contacting the individual. This is even more important if you need to present the information in a legal context. Therefore, you need an unaltered rendition of the original audit log.

However, you also want a normalized view of the data so that complex analytics aren’t dependent on nuances of the log files of each source vendor. HawkEye AP accomplishes this with a common database construct known as Views as well as an extension to the SQL language called TableMatch. The TableMatch functionality finds all the related tables automatically. This means you can add new sources and the data will show up in reports (and advanced analytics) without any code changes. All together this is our trademarked IntelliSchema solution. Since Views normalize the data at query-time, not at load-time, the original raw data is preserved while queries can access the data in a generic manner.

Need Volume Capabilities to Establish Baseline
Huge amounts of data are needed to establish baselines. This is clearly a big data problem. Some of the statistics of current HawkEye AP production implementations include:

- Multiple Petabyte+ customers, dozens of customers over 100 Terabytes
- Loading over 7 billion records daily with peak loads over 20 billion records in a day (without interruption to reporting or other activities)
- Scanning trillions of records daily
- Loading data from 20,000 Windows Servers daily
- Loading data from 1,300 MS SQL Servers daily
- Loading data from 1,500 IIS web sites deployed on 40+ IIS Servers daily

You need to make sure your log management system can scale for both your current and future data needs.

3. Execute the Data Flow Model

Execute Aggregation Queries to Calculate Metrics
The model begins the simple “Execute Report” component that runs a standard summary report over normalized views of log data for each different behavior. These reports are also used in daily reports for management, auditors, or interested parties to understand their organizations ongoing behavior. The reports will also be used as drill-down targets for inspection of outliers identified in the insider threat executive dashboard.
Using HawkEye AP 6.0 for Detecting Insider Threats

Maintain Metrics in Rolling Summary Table

The next component in the model saves each day’s summary information into a rolling table of metrics. This table will be used to calculate the statistical average and standard deviation, and provides a history of each individual’s scores over time. This data is stored in a standard read/write database such as PostgreSQL. Taking advantage of the ability to nest models inside of other models, all the operations of maintaining the table can be accomplished in one simple component. Logic is included to remove old rows and to detect which days are missing from the metrics table so it is not overpopulated if the model is executed more frequently than the schedule.

Recursively Calculate Score Across Hierarchy of Behaviors

Once the information is in the metrics summary table, a process can “climb the tree” of behaviors and calculate the average score and standard deviation for the whole community for each behavior. It can then iterate through the list and calculate the historical average and standard deviation for each individual, which are additional key components for insider threat risk-ranking and detection.

Generate Report Data

Finally, the model can store the results in a table for reporting. Data modeling is independent of reporting or presenting the information so that organizations can either use the HawkEye AP reporting and dashboarding infrastructure or leverage any standard BI tool to visualize the information.

Scheduling the Model

Once the model is created it can be scheduled just like scheduling a routine report. By default, the model is designed to be run daily, off business hours. The model can easily be scheduled to run less frequently without needing any adjustments to the reports or models. With minor adjustments to the reports the models can be scheduled to run more frequently.
4. Visualize the Results

Now we get to go back to the reporting and dashboard features of HawkEye AP and see the results. The insider threat detection solution comes complete with a configurable Executive Insider Threat Dashboard for displaying the most important charts and grids from the insider reports. The dashboard comes as shown; however, content can easily be added, removed, resized, or re-positioned on the screen. Further, the grids and charts can be re-used on any number of additional dashboards that can be created and configured just as easily.

The first grid on the dashboard lists every user in the organization ordered by their risk indicator score, labeled "IT-score" in the grid. The next column is the "Z-score" which is the statistical description of how much each user’s score deviates from the average score of all employees (number of standard deviations from the average). Users can quickly zoom into a full-screen view of just this report to see each user’s IT-score from the current run, or their 7-day average, 30-day average, 90-day average, etc. From this view users can drill-down into a report that shows the makeup of a single individual’s IT-score.

The next chart on the dashboard is a graphical view of the historical trend of three highest risk individuals, in context with the average risk of all employees and a plotted line showing a basic three-standard deviation threshold. The horizontal axis represents time, so that you can see how many times an individual has crossed the threshold over the entire time period (which is programmable). When viewing this chart, users can interact with it by using the mouse to quickly zoom into a specific time-frame, or click on labels to eliminate unwanted information.
Using HawkEye AP 6.0 for Detecting Insider Threats

Additional charts plotting the Z-score depict the highest risk users with the vertical axis showing the number of standard deviations away from the average. In this case the average will be a horizontal line as it is always 0 standard deviations from itself. Z-score plots such as this, allow an analyst to visualize deviations from the community average, and deviations from the user historical average.

Conclusion

This insider threat detection model is just one example of what big data and advanced analytics can do for security. This model looks only at exfiltration to find insiders; two other strategies are to look at sabotage and fraud. Hexis Cyber Solutions is currently developing models for these other two strategies as well as more advanced visualization techniques. Still all of insider threat detection is just scratching the surface of what can be done with these data flow models.

About Hexis Cyber Solutions

Hexis Cyber Solutions (Hexis), a wholly-owned subsidiary of The KEYW Holding Corporation (NASDAQ: KEYW) based in Hanover, Maryland, provides complete cybersecurity solutions for commercial companies, and government agencies.

Our mission is to ensure that business IT infrastructure is equipped with tools and capabilities to detect, engage, and remove both external and internal cyber threats. Cyber terrorists, organized crime, and foreign governments focus tremendous effort on commercial, government, and military interests as their prime targets. Hexis’ HawkEye and NetBeat products offer active, multi-disciplined approaches to achieve a higher standard of cybersecurity that is based on our expertise supporting cybersecurity within the US. Visit www.hexiscyber.com.