Wavecrest

UNDERSTANDING BEHAVIORAL ANALYTICS WITH RELIABLE METRICS

The

HUMAN FACTOR

White Paper Series

www.wavecrest.net

In managing the human factor in the workplace, employers must understand employee behavior and the data employees are creating. In addition, employers must monitor the flow of the data in and out of the organization. By doing so, they can detect patterns of human behavior that could indicate insider threats, inappropriate or excessive Internet use, or events that could lead to a data breach. Creating an audit trail is also essential to monitor employee behavior. This paper is one in a series of papers that addresses the human factor of data security using behavioral analytics focused on human behavioral patterns, and interpreted with relevant report metrics.

Introduction

In your organization, do you suspect that an employee is visiting common job search sites or accessing personal cloud storage sites? Maybe it's a disgruntled employee who is dissatisfied with his or her job. Could there be a data breach occurring? Research indicates many significant data breaches are ultimately an "inside job." Insiders could be employees, contractors, business associates, or partners–humans–who pose the biggest risk to enterprise data, since they are allowed access to sensitive data. When an employee exhibits a pattern of behavior that includes visiting popular job sites, such as Indeed, LinkedIn, or Monster, the employee may be planning to leave the company. Perhaps the employee is also uploading confidential information to personal storage sites, such as Google Drive, Dropbox, and iCloud.

Enter behavioral analytics. Behavioral analytics focus on patterns of human behavior allowing management to understand what is normal and flagging anomalies that indicate insider threat. With employee Web-use behavioral analytics, the organization can set up a solution that ties Web requests to employees and performs Web categorization of these requests delivered in logging. Using this log data, report metrics in the analytics reporting feature assist IT, HR managers, and other department heads with identifying potential insider threat behavior, assessing trends for data breach exposure, and observing human behavioral patterns for lost productivity. Interpreting behavioral analytics with reliable metrics also exposes abnormalities in user activity and flags possible legal liability issues.

There are many metrics available in behavioral analytics reporting, such as hits, visits, time online, bytes, download time, and denied visits, and more than one metric is necessary to provide meaningful insight into employee behavior to managers. The most important metric by far is visits that gauge the level of employee Web activity, but it is often confused with hits. Visits give you uncluttered, relevant Web activity detail based on user clicks, whereas hits consist of unsolicited traffic and are not a reliable way to measure employee Web site traffic.

It is important to note that most behavioral analytics solutions only provide hits and do not engage in visits analysis which is critical to understanding human behavior. Visits are not inherently available in the raw data in Web traffic. They are derived from a unique algorithm, which if not accurate, can also result in incorrect data for other metrics such as time online. In addition, if visits are not calculated correctly, managers reading Web activity reports will have no understanding of actual user activity and will be in a difficult position to address any behavioral issues with employees. The ability to determine visits with accuracy affects the quality of an employee Web activity analysis and may or may not be offered by an analytics solution provider. Given that today's Web sites are much more dynamic and complex than many years ago, separating hits from visits is even more challenging.

When a user goes to a site or URL, that action can be thought of as a single, clickable, manual event. But as soon as the user clicks the mouse or presses Enter, a myriad of elements are loaded in the browser. These can include text within the HTML file, text pulled from a content delivery network (CDN), images, videos, cascading style sheets (CSS), and JavaScript. Because each of these arrivals has its

own URL, all of them–those that were triggered manually and those that were triggered automatically–are individually recorded at the server location. Such triggered log records, known as log files, make no distinction between manual and automatic events. As you can see, one click by one user can result in the logging of many events, often referred to collectively as activity. Does all of this manual and automatic Web activity constitute hits, visits, or both? To answer this, a further explanation of these two metrics is provided below.

An explanation of hits versus visits

Confusion with hits and visits may lie in the fact that a visit is also a hit, and hits produce larger counts. Both hits and visits should be determined in the behavioral analytics reporting feature. The following explanation distinguishes between the two and describes how they are meaningful.

Hits

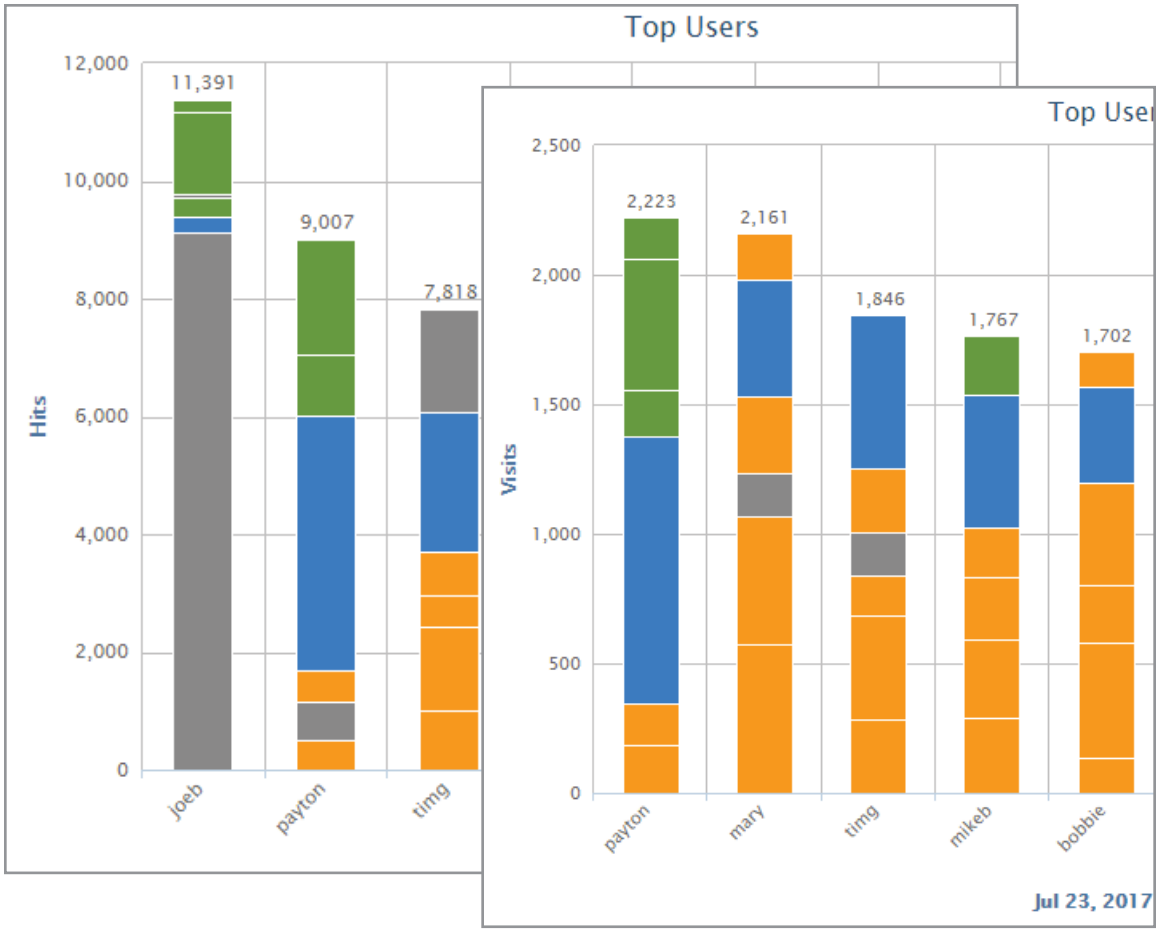
A hit is any request to a Web server. Each time an employee loads a Web page, clicks a hyperlink, views a graphic, or performs any other action on a Web site, a call is made to the Web server. The Web server records each of these requests in a log file. These requests are commonly known as hits, and the loading of a single Web page can amount to many hits, due to all of the elements it contains. Images, JavaScript, cascading style sheets, embedded objects, and other Web site elements all contribute to the Hits count.

In other words, a hit can be defined as any visit to a Web site, plus any multimedia images and advertisements. Hits represent the number of requests for Web page elements that the Web site fulfilled. For example, a single page with 30 images, 1 JavaScript file, and 1 style sheet would be 33 hits, including the page as a hit in the total count, that is, 1 visit, but 33 hits. A hit can be a meaningful measure of how much traffic a server can handle.

Visits

A visit is a click action for the purpose of visiting a Web site. One click equals one request for a Web page. A visit does not include multimedia URLs, such as graphics or audio pages, banners and advertisements, or Web pages that were requested as part of a visit, that is, unsolicited. For most accuracy, behavioral analytics could provide the ability to set a reasonable time period so that the same URL visits are not counted more than once in reports in that time period. For example, if the time period is set to 3 seconds, all requests in this time period are considered hits, and the URL is counted as only 1 visit.

Visits represent the number of times a Web site was accessed. A visit is a meaningful indicator of the level and type of Web activity occurring in your network. In analytics reporting, visits can be delivered in several different perspectives, such as visits by category including legal liability and cloud service categories, by classification rating such as acceptable and unacceptable, and most importantly, by user.
In the example below, the Top Users charts show the top 10 users with the most hits and visits for a specific time frame. Notice that the top user changes depending on whether the metric is Hits or Visits, and the Hits count is much larger than the Visits count for the same top user, for example, payton.



Hits Versus Visits

Additional metrics aid in understanding human behavioral patterns

As mentioned above, behavioral analytics rely on more than one metric to accurately interpret employee behavior. In addition to hits and visits, other reliable report metrics, including time online, bytes, download time, and denied visits, are necessary to properly analyze human behavior. Briefly described below, these additional reliable metrics, along with hits and visits, deliver the most accurate results in Web activity reporting.

Time Online

To gain an understanding of time online, a description of "browse session" follows. Opening a browser generates Web traffic. This represents the beginning of a browse session. A session is a group of interactions that take place on a Web site within a given time frame. For example, a single session can contain multiple screen or page views, events, social interactions, and e-commerce transactions. The session is open as long as Web traffic is continually being generated. The session is considered closed once an amount of time passes with no Web traffic. A new browse session begins as soon as Web traffic is generated again.

Time online is an approximation of the time that a user spends on the Internet, based on the time stamps from Internet requests made as the user browses Web sites, the average number of minutes for reading a specific Web site, and the time spent reading the last Web site before the end of the browse session.

Bytes

The Bytes metric is used as a measure of bandwidth consumption and may be displayed in kilobytes, megabytes, gigabytes, etc. This metric allows you to view the bandwidth for your top consumers, that is, users and groups who are consuming the most bandwidth, as well as the bandwidth used by the top content categories, acceptability classifications, and Web sites. You can detect unexpected spikes that could indicate excessive bandwidth or Web use.

When interpreting behavioral analytics, you can also get the number of bytes read per hour, per top user, per top bandwidth site, and per Web request, and much more.

Download Time

Download time is the approximate or average time for a Web page to load in the browser, that is, the period between the time that a user clicks a hyperlink and the time that the page loads in the browser. In analytics reports, the Download Time metric can be derived by multiplying the smallest average amount of time required to download a typical Web page by the number of visits. For example, if the smallest amount of time to download a Web page is set to 3 seconds, and 100 Web pages were loaded, 100 multiplied by 3 seconds each = 300 seconds of estimated download time. Download time is usually displayed in a DD:HH:MM:SS format, where DD=days, HH=hours, MM=minutes, and SS=seconds, in reports.

Denied Visits

Another report metric is the Denied Visits metric which indicates denied requests for a Web page. There are many reasons why a user may be denied access to a Web page. These include that the user may not be authorized to receive the page, the page may not have been found by the Web server, or the page may have been blocked for access. An analytics report dedicated to denied visits can be used to identify users who may be engaging in excessive attempts to visit inappropriate or unauthorized sites.

Conclusion

Behavioral analytics seek to answer the human factor questions: Which users visited which sites? When did they do so? How often did they do so? What type of content were they seeking? How much bandwidth was consumed in the process? Were the visits in compliance with the organization's Acceptable Use Policy? With the necessary metrics, organizations are able to analyze employee behavior and answer these questions. A behavioral analytics solution with reliable metrics is a must to protect your valuable assets, to spot insider threats, and to detect events that could lead to a data breach.

Metrics, such as visits, hits, time online, and download time, quickly give managers the detailed data that they need to get insight into how employees are using network resources and help address the human factor in the workplace. Activities, such as application usage in the middle of the night or on weekends, unusually large file transfers, excessive online searches, and more, can all be identified and flagged as abnormal behavior. However, managing employee Web use, that is, the process of ensuring employees use Web access appropriately, is not necessarily an IT issue. The next paper in the series will discuss the people issue of Web-use management and what constitutes an effective employee Web-use management program.

About Wavecrest Computing

Wavecrest has over 20 years of proven history of providing reliable, accurate Web-use management and Advanced Log File Analyzer products across various industries. Managed Service Providers, IT Specialists, HR professionals, Forensics Investigators, and business managers trust Wavecrest's Cyfin and CyBlock products to manage the human factor in business Internet usage–managing cloud services, reducing liability risks, improving productivity, saving bandwidth, and controlling costs. Wavecrest is trusted by large government and commercial organizations such as US-CERT Homeland Security, U.S. Department of Justice, USPS Office of Inspector General, National Grid, Johns Hopkins, and a growing list of global enterprises and government agencies. We are a proud long-term GSA contract holder. For more information on the company, products, and partners, visit https://www.wavecrest.net.

Wavecrest Computing
904 East New Haven Avenue
Melbourne, FL 32901
toll-free: 877-442-9346
voice: 321-953-5351
fax: 321-953-5350