Countering Cyber Threats by Modeling "Normal" Website Behavior

A SOLUTION THAT LEVERAGES BIG DATA COLLECTION, STATISTICAL MODELING AND ABNORMAL-PATTERN IN REAL TIME CAN FOIL ATTACKS AIMED AT EXPLOITING BUSINESS LOGIC FLAWS.

More than three decades after futurist John Naisbitt published his seminal book Megatrends, one of his most famous observations has never seemed more prescient: "We are drowning in information, but starved for knowledge," he wrote. Today's flood of digital data—flowing from hundreds of millions of devices and contained in billions of online transmissions and transactions—often presents organizations with more raw information than they're able to absorb. Even more worrisome, this data deluge provides cyber criminals with a means to hide their malicious activities within the vast ocean of legitimate digital streams.

Having lots of data doesn't have to be a bad thing. Indeed, most IT and business executives have come to recognize there is great value to be mined from the huge volumes of data constantly being generated by their employees, customers, market researchers and others. Consequently, "big data" collection and analysis has emerged as one of the most promising IT trends, allowing companies to tease out useful business knowledge from the pervasive data "noise."

Many of these same executives, however, don't know that statistical modeling and analysis based on big data can also help them counter a fast-emerging breed of cyber threats as they're happening. Too few organizations, in fact, fully grasp the nature of these threats, which seek to exploit the very business logic on which their business processes and websites run.

The fact that many organizations may have gaping holes in their IT security armor may come as a

RSA CSO Strategic Marketing Services

surprise. After all, it's no secret that corporate data and applications represent the lifeblood of most businesses, and companies collectively spend billions of dollars each year to protect these digital resources. IT security has gained an even higher profile of late following the rise of cloud computing and mobility among employees. Both of these trends can deliver huge benefits, but they also introduce new types of security risks to the organizations seeking to capitalize on them.

Indeed, the same characteristics that make cloud- and mobile-based interactions simple and attractive to consumers and employees also open wide avenues for cyber criminals to roam. Compounding this problem is the sheer volume of data now traversing the digital network. For example, according to network equipment vendor Cisco, global traffic on mobile networks alone totaled nearly 18 exabytes in 2013. During the same year, ecommerce sales in the U.S. totaled $263 billion, reports the U.S. Department of Commerce. That represents nearly a 17 percent increase from the nation's 2012 ecommerce sales.

Faced with skyrocketing digital traffic and an ever-expanding threat matrix, companies have doubled down on sophisticated identity management solutions, installed powerful firewalls, invested in advanced data encryption methods and deployed a wide range of other protection measures. While necessary and largely effective, however, these technical barriers can sometimes be breached. If they are, organizations can now turn to a solution driven by big data: modeling normal online behavior of a company's website users, and instantly flagging actions that fall outside of those established norms.

THE DARK SIDE OF A DATA-DRIVEN, ONLINE WORLD

Fewer than 20 years after the emergence of the commercial Internet, it's impossible to imagine a world without ecommerce, online banking and thousands of other forms of Web-based business interactions. A company's Web presence is often as or more important than its physical brick-and-mortar sites, assuming it even has such sites. The dividing line between computing operations and business operations has faded to the point of invisibility, with many business processes inseparable from, or impossible without, the digital foundation on which they run.

With the number and value of online transactions skyrocketing, cyber criminals are finding it easy to hide their activities alongside those of legitimate Web and mobile users. In a 2012 study conducted by the Ponemon Institute and Silver Tail Systems (since acquired by EMC/RSA), 74 percent of security professionals surveyed said they had difficulty telling the difference between actual customers and criminals accessing their websites. Nearly as many, 69 percent, said they didn't have the proper tools in place to counter business logic abuse, according to a subsequent study conducted by the Ponemon Institute and RSA Security.

So what, exactly, is business logic abuse? This type of online criminal activity and fraud can take many forms, but all share a common trait. The attacks seek to disguise their malicious intent by cloaking themselves within the normal operations and functionality of a website. If successful, such business logic attacks can bypass, or even enlist, existing security controls—all while masquerading as just another benign customer interaction.

THE SAME CHARACTERISTICS THAT MAKE CLOUD- AND MOBILE-BASED INTERACTIONS SIMPLE AND ATTRACTIVE TO CONSUMERS AND EMPLOYEES ALSO OPEN WIDE AVENUES FOR CYBER CRIMINALS TO ROAM.

Consider, for example, a website that has instituted an authentication rule giving a user five chances to enter the correct username and password. After five failed attempts, the user is automatically locked out from accessing the website. But what if this rule is set up to cache the logon activity on the browser, rather than on the server? In that case, notes Eric Thompson, IT security strategist at EMC's RSA security division, the attacker could repeatedly close and restart the browser to gain innumerable chances to guess legitimate logon credentials. Failing to recognize this business logic flaw, the website operator might assume that its five-chances authentication rule was effectively stopping unauthorized users from repeated logon attempts.

Some business logic attacks can occur, as in the password-guessing example above, even before an attacker gains access to a site. One of the most common of such attacks is the distributed denial-of-service (DDoS) assault, which has become all too familiar to banks and other financial institutions. DDoS attacks enlist hundreds or thousands of

computing devices that have been compromised with some form of malware. An attacker, often a robotic "botnet," then directs the compromised devices to flood a target website with high volumes of traffic. At a minimum, a DDoS attack can slow a website or knock it entirely offline. But these attacks can also serve as cover to launch other attacks, such as account takeovers or fraudulent money transfers, while the site's operators are preoccupied with countering the deluge of DDoS traffic.

Financial institutions are far from the only targets of business logic attacks. Ecommerce sites can also succumb to password-guessing and DDoS assaults, but have their own forms of cyber crime to counter as well. One example observed by RSA involved an online marketplace being bilked by criminals who had registered on the site and were then making phony purchases using real credit cards. They then claimed a 10 percent rebate that the vendor offered on all purchases, pocketing more than half-a-million dollars in rebate payments before the marketplace deployed RSA's technology, which uncovered the scam.

Cyber criminals in this case and others have been smart enough to lay the groundwork for such attacks over time. "You don't want to launch an attack from an account that's one-hour old," notes Jason Sloderbeck, RSA's director of product management. Attackers will often mass register a large number of accounts on a site over an extended period, he says, and will wait to use them for an attack to avoid setting off any red flags.

Business logic attacks of this type and many others often can't be caught or countered by existing security measures, but the resulting losses, in both monetary and reputational terms, can be staggering. (For examples of some of the many forms of business logic attacks, see "Cyber Attacks Grow More Diverse.")

ONLINE FRAUD CONSEQUENCES GROW, SPURRED BY THE RISE OF MOBILITY

Financial institutions and ecommerce sites represent some of the most tempting crime and fraud targets, but no organization with an online presence is immune to cyber threats. The loss in business opportunity can be significant if an attack brings down a company's website, and criminal access to sensitive corporate information can have severe monetary as well as legal ramifications. More difficult to quantify is the negative impact on an organization's reputation if it and, especially, its customers are victimized by a successful attack.

In one measure of the potential consequences, the Ponemon Institute and RSA study calculated the impact of downtime on Cyber Monday, the post-Thanksgiving kick-off of the online holiday shopping season. The study's authors found that a single hour of downtime could cost U.S. retailers up to $3.4 million in losses associated with brand damage and reduced consumer confidence.

In another analysis—this one conducted by RSA and the Aite Group—global losses resulting from account takeovers totaled about $454 million in 2012. The study's authors projected such losses will grow to $794 million in 2016 as they span multiple attack vectors.

MORE DIFFICULT TO QUANTIFY IS THE NEGATIVE IMPACT ON AN ORGANIZATION'S REPUTATION IF IT AND, ESPECIALLY, ITS CUSTOMERS ARE VICTIMIZED BY A SUCCESSFUL ATTACK.

Within the spectrum of attack vectors, few are more challenging than attacks originating from smartphones and other mobile devices. More than 1 billion smartphones shipped in 2013 according to market research firm IDC, and these mobile devices have become the default Web-access platforms for many consumers and employees. In 2013, U.S. retail mobile commerce sales already totaled $42.13 billion, or 16 percent of the total $263 billion in U.S. ecommerce sales, according to market research firm eMarketer. By 2018, the research firm predicts, those mobile sales will total $132.69 billion, representing 27 percent of the projected $491.44 billion U.S. ecommerce market that year.

The rise of mobility presents website operators with a quandary, and cyber criminals with an attractive new mode of attack. Compared with PCs, smartphones and other mobile devices are relatively difficult to identify by security methods that attempt to authenticate individual devices. The mobile devices also use different protocols than PCs, and they routinely encrypt the data they're storing and transmitting.

"With mobile, everything is opaque," says RSA's Sloderbeck. "We've seen cases where the bad guys

realize mobile is different. They know if they find the mobile API and hit it, it won't trigger the same kind of protections as on the Web environment."

Indeed, in tracking the activity of one group of its customers, RSA found that although just 27 percent of the transactions originated from mobile devices, 32 percent of those mobile transactions were fraudulent.

LEVERAGING BIG DATA AND STATISTICAL MODELING FOR EXTENDED PROTECTION

For beleaguered banks, online retailers and other Web-based businesses, the scope of the cyber threat can seem overwhelming. After all, the number and variety of potential threats rooted in business logic are all but limitless, the volume of online transactional traffic is on a hockey-stick growth curve, and the number of Web-connected devices of all types already totals in the billions. At the same time, most organizations are increasingly—and sometimes fully—dependent on online business models of one sort or another.

In something of a serendipitous twist, however, the big data traffic volumes that can mask fraudulent cyber threats can also be employed to expose them. Companies already have the means to track individual Web-session click streams in real time, but that capability alone can't proactively identify potential threats among the millions of activities that may be occurring simultaneously on their sites. What's needed is visibility into what is normal and nonthreatening behavior on any given website. Only with that visibility can organizations hope to immediately spot suspicious behavior that falls outside of the norm.

THE BIG DATA TRAFFIC VOLUMES THAT CAN MASK FRAUDULENT CYBER THREATS CAN ALSO BE EMPLOYED TO EXPOSE THEM.

Determining normal Web-based behavior may seem a straightforward task at first blush, but it is far from being so. Consider that each website has its own business processes, its own base of users and its own security controls and policies. Each site is also likely to experience a huge range of activities and website navigation that fall within the "normal" range of its operation.

"You need to look at the sequence of pages accessed, the velocity of movement across the site, what parameters are often submitted to a page [and] what is never submitted during the user's session," explains Sloderbeck. Only by tracking and analyzing these and hundreds of additional variables can companies hope to create statistical models that accurately depict the range of interactions common to their unique websites.

Described another way, this type of big data analysis and Web-user profiling makes use of crowd-based analytics. "Once you have visibility into the entire crowd, you can identify normal patterns and whitelist them," says RSA's Thompson. "From those normal patterns, you can then identify deviations." Cyber criminals might try to fit within expected patterns, for example, by programming a botnet attack to click page links at random intervals in an attempt to appear humanlike. "Even if they programmed in 'random,' though, that randomness itself would be abnormal," Thompson says.

REINING IN BUSINESS LOGIC ATTACKS

To counter business logic–based cyber threats, organizations need to use big data "crowd" analytics, sophisticated statistical modeling and deviant-pattern recognition for an additional layer of protection on top of their existing security measures. It's important to identify behavior-based threats that other controls can't spot, and work hand-in-hand with those controls to deliver a more nuanced and proactive approach to protecting a company's assets and reputation.

Big data–based threat detection solutions can ship with built-in rules to immediately detect suspicious traffic as soon as they begin monitoring a site's traffic. Ideally, however, the solutions will then "learn" what is normal behavior for a specific website, and will continue to fine-tune their models. With the proper sampling and modeling algorithms and access to a website's traffic, an accurate statistical model emerges of what behavior constitutes common and expected activity—ideally, using streaming analytics to update statistical models in real time with every single click.

In a world awash in Web traffic, scalability is a core requirement of any security solution that hopes to tame big data. The need to field a highly scalable system is one reason organizations may decide to purchase a cyber crime–detection solution rather than attempt to develop one on their own.

Building a solution that can monitor all the Web traffic of a large organization is no trivial task, notes RSA security strategist Thompson. Furthermore, he says, the solution should also provide an easy-to-use interface for monitoring traffic and responding to threats. Another key consideration in the build-versus-buy decision is speed to market, where organizations must determine the trade-offs of designing, building and testing a custom-built solution compared with purchasing a commercially available product.

Most fundamentally, notes RSA's Sloderbeck, is the level of expertise required to build a sophisticated, real-time Web threat detection system. "We have PhDs in math building out our behavioral models," he says. "Companies thinking of taking on this challenge need to ask themselves if this is really their core competency."

EMC/RSA is putting its expertise to work, leading innovation in the diverse and ever-evolving security landscape. RSA Web Threat Detection, the company's big data solution for identifying and countering abnormal website activities, is currently protecting more than 10 billion Web requests per day for RSA's customers, Sloderbeck says. "We're protecting 40 percent of the North American banking traffic, and are analyzing the traffic of more than 1 billion Web users each year." And, because RSA Web Threat Detection continually monitors what is "normal" website behavior and fine-tunes its models accordingly, the solution will be able to counter future forms of business logic attacks as well as the wide range of attacks already occurring today.

CYBER ATTACKS GROW MORE DIVERSE

Cyber criminals seeking to exploit business logic vulnerabilities have no shortage of opportunities, but ecommerce and banking/finance sites are favorite targets. Some of the most common attacks are:

ECOMMERCE SITES

- **Vulnerability probing**—exploring a site's operations and business logic to identify possible entry points
- **Password guessing**—running robotic attempts at guessing legitimate user ID and password credentials, sometimes by subverting existing measures designed to limit failed attempts
- **Promotional abuse**—exploiting flaws in sales and rebate programs
- **Stolen credit card testing**—making purchases with stolen credit cards to verify that they are active
- **Site and inventory scraping**—copying the content of a site, typically to set up a fake site to capture customer IDs and credit card numbers

BANKING/FINANCE SITES

- **New account fraud**—the theft and use of personal information to open new credit card accounts or other accounts without the user's knowledge
- **Automated clearing house (ACH) fraud**—a variety of techniques to steal funds from the ACH network, the central clearing facility for electronic funds transfer transactions in the U.S.
- **Check image scraping**—a technique for scraping check information from a banking site, presumably to use in fraudulent activity
- **Man-in-the-browser**—a proxy Trojan horse that infects a Web browser in order to modify Web pages or transaction content, or to initiate additional transactions
- **Man-in-the-middle**—a form of eavesdropping in which the attacker makes independent connections between victims, and relays messages between them, sometimes inserting new messages from the attacker

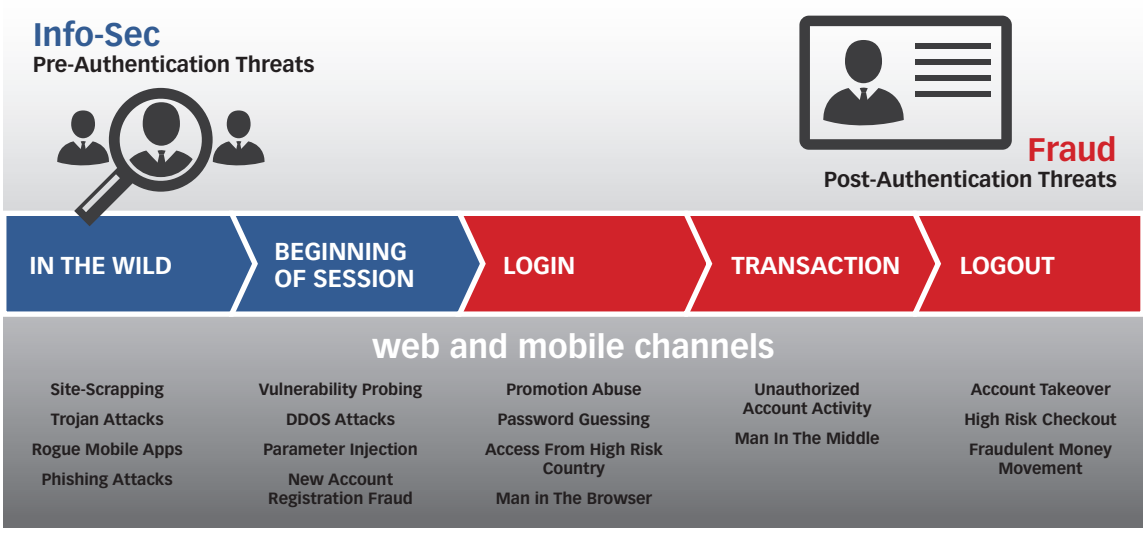
Detecting Sophisticated Online Attacks with RSA Web Threat Detection

Security professionals are aware that cyber criminals have increasingly sophisticated weapons at their disposal for maneuvering through online commerce systems and stealing information. Traditional firewalls, IPS/IDS, and web application firewalls (WAFs) do little to help online businesses understand the behavior of website visitors. Instead, they narrowly focus on the network and server exploits only.

Because of this gap in technology, cybercriminals are evolving to exploit a new attack vector known as business logic abuse, which results from criminals exploiting flaws in the business functionality of a website to take over user accounts, steal money, scrape information and perpetrate other types of fraud. These costly threats are growing significantly, making the job of IT professionals extremely complicated.

THE TRADITIONAL APPROACH

Existing solutions for detecting and analyzing online criminals behavior usually identify either pre-authentication threats (infosec products) or post-authentication threats (fraud products) – but not both.



Prevention

Armed with knowledge of the types of attacks their sites may face, IT professionals can use firewalls or rewrite software to block the traffic they assume to be dangerous. However, the business logic abuses favored by today's cybercriminals are beyond the scope of preventive solutions. Fraudulent activity committed by abusing shopping cart logic, or on a 3rdparty bill-pay site, happens alongside legitimate traffic.

Detection

The challenge of detecting anomalous activity in real-time requires gathering various "big data" sources and correlating them to understand user behavior. However, current methods of detection fall short of this goal – individually, they examine only pieces of the behavior puzzle, not the entire picture. Web application firewalls (WAFs) can examine transaction signatures, but will only block traffic previously identified as a potential threat. Security information and event management (SIEM) solutions use limited data in log files to seek out behavior that seems anomalous. These solutions can only identify broad trends or rule violations; they can't correlate anomalies to individual user sessions.

Investigation

Flagging certain transactions for review is a commonly used tactic for detecting fraudulent activity – and it served its purpose in the days when cybercriminals used exploits such as brute force attacks or stolen credit cards. However, the data that information security teams need to identify suspect transactions is often scattered among multiple systems. For example, website logs are with web admins, IP level information is with the network team, and account information is with

THE CHALLENGE OF DETECTING ANOMALOUS ACTIVITY IN REAL-TIME REQUIRES GATHERING VARIOUS "BIG DATA" SOURCES AND CORRELATING THEM TO UNDERSTAND USER BEHAVIOR.

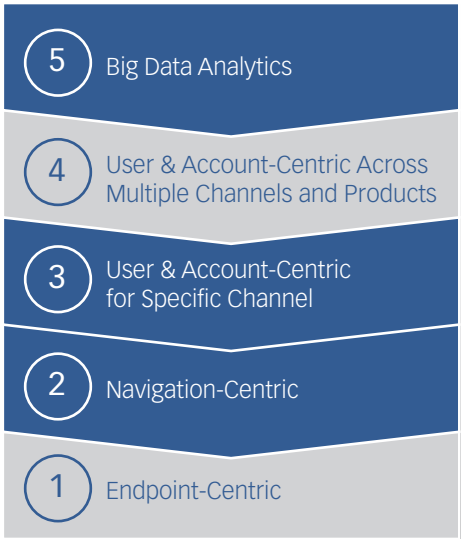
operations. And since the team can't see a user's specific session on the website, they can't compare the incidents they are flagging to the actual use activities, so the investigators gain no insights on behavior. A more centralized approach that allows atomic analysis is needed.

IDENTIFYING CRIMINAL BEHAVIOR ONLINE

When you have so many people interacting with your website on a daily basis it can be difficult to tell the difference between legitimate and criminal users – after all it is virtually impossible to monitor what every individual is doing at all times.

Cybercriminals exploit this lack of visibility into user behavior by hiding themselves and their activities among legitimate users and legitimate activities – making it extremely difficult for organizations to detect these types of attacks in real time. Rather, they must rely on log and other retrospective data to investigate the cause after an attack has become a reality.

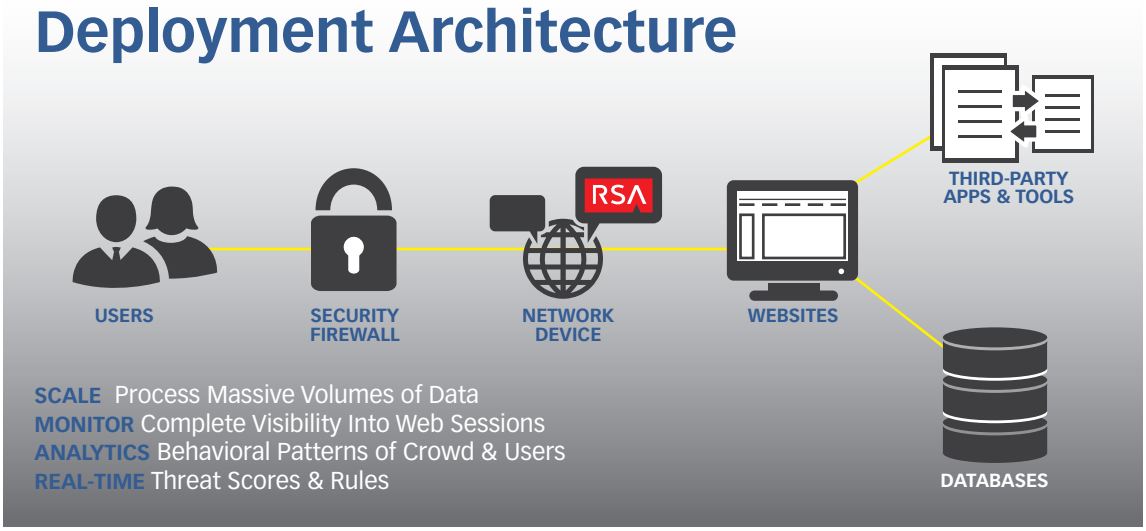
This results in low fraud detection rates, high costs of manual review and increased exposure to threats. RSA Web Threat Detection helps identify potentially criminal use of a website by detecting anomalous online behavior – behavior that is out of the ordinary



The Five Layers of Fraud Prevention

"No single layer of fraud prevention or authentication is enough to keep determined fraudsters out of enterprise systems," according to Avivah Litan, vice president and distinguished analyst at Gartner. "Multiple layers must be employed to defend against today's attacks and those that have yet to appear."

RSA Web Threat Detection is unique in its ability to detect anomalous website behavior by utilizing layers 2, 3, and 5, all in one platform.



from general population of web visitors. This allows the information security and fraud teams to focus their attention on the users that have exposed themselves as potentially disruptive rather than trying to identify the cyber equivalent of a needle in a haystack.

INNOVATIVE AND EFFECTIVE ONLINE THREAT DETECTION

One of the key differentiators of the RSA Web Threat Detection solution compared to other web analytics suites is its ability to "sessionize" a user's clickstream. What this means is that every click a user makes on a website to navigate from login to checkout/logout is grouped together so that the entire stream can be utilized for analysis on a user to crowd, user to user, and individual stream basis. This is achieved in real-time by a number of innovative components.

Sessionization of Data

RSA Web Threat Detection is software installed on a server in the data center, which sniffs packets from a SPAN port configured to mirror traffic from a web server. Architecting the deployment is this manner allows Web Threat Detection to be completely separate from the web server and have zero impact on latency or increased risk for connectivity issues for the end user.

THE RSA WEB THREAT DETECTION SCORING ENGINE BUILDS STATISTICAL MODELS BASED ON CROWD BEHAVIOR, WHICH REPRESENTS THE GOOD BEHAVIOR.

Web Threat detection then filters and reassembles the packets to extract the TCP payload. It then parses any of several protocols, including HTTP and HTTPS, in order to extract important attributes and create metrics about the web traffic at all levels of the protocol stack. Once this is all completed, the resulting messages are distributed to other core elements of Web Threat Detection performing tasks like logging, scoring and reporting.

This allows fraud analysts to run analysis or rules on any piece of data pertaining to the traffic on their website, and for the automated Scoring Engine to run its bleeding-edge Streaming Analytics upon it as well, in real-time.

Threat Scoring

The RSA Web Threat Detection scoring engine builds statistical models based on crowd behavior, which represents the good behavior. Each user session is then compared to this model and analyzed as to whether it fits the good behavior or falls under anomalous behavior. This real-time Streaming Analytics is performed in real-time on a click-by-click basis. Further, this scoring engine is able to store an in-depth traffic profile composed of hundreds or attributes in memory so that the following scores can be calculated for each click:

- **Velocity:** These scores measure the speed of each page transition across the population of all visitors on the site and then measuring the divergence of a particular visitor from those norms.
- **Behavior:** These scores identify unusually frequent behavior. It measures how far a session's

navigational sequence is different from the norm across all sessions.

- **Parameter:** If a parameter is submitted that is rare, the engine is able to detect these anomalies and score based upon them.
- **Profile:** These scores follow the same logic as the behavior scores. However, these are compared against an individual user's history over a period of time.
- **Periodicity:** The periodicity score provides pattern recognition of time between clicks which helps detect robotic activity.
- **Man-in-the-Browser (MiB):** MiB attacks are performed by malicious code operating within the user's browser or on the user's machine and operate within a time span of a user session. A user's page transition patterns, geo IP, speed of actions, alongside multiple other elements determine the score.
- **Man-in-the-Middle (MiM):** MiM attacks are characterized by situations in which the attacker has gained access to, and is exploiting, the web session of another user. Once this pattern has been identified, a series of discounts are made to determine the validity of a MiM attack, these include user agent, IP address, transfer amount and page transitions.

The Decaying Data Model

The baseline data used while generating scores is organized into Data Models. A data model consists mainly of frequencies of tracked items over a defined period of time. Decay factor is then configured, which works by setting an amount of time for mathematical degradation of the data's strength for comparison over time. This is done so that the most recent data is given the most strength.

The best practice for defining both attributes depends on the computational resources available and the dynamic characteristic of the site. Once a counter's decay has reached 0, its record is pruned to prevent unbound growth of obsolete information.

The Analysis Window

The analysis window is a sliding window of time representing the portion of traffic to be analyzed when compared with the current data model for scoring threats. In real-time usage, the analysis window ends at the present moment. Session scores are calculated as the sum of all transaction scores within the analysis window. To ensure that we are comparing the counters from within a session to each individual session, this window is configured as close to the average session time of a user on the website as possible.

THE ANALYSIS WINDOW IS A SLIDING WINDOW OF TIME REPRESENTING THE PORTION OF TRAFFIC TO BE ANALYZED WHEN COMPARED WITH THE CURRENT DATA MODEL FOR SCORING THREATS.

The model frame and the analysis window are largely independent. The model frame determines the population data that is used to calculate transaction scores within the analysis window. Changes to the data model due to the decay of data in the model frame will have little effect on the analysis window calculations. However, requirements for the time periods of the model frame, analysis window and sessions are configured to average out the hourly, daily and weekly business cycles of the website.

TRADITIONAL SOLUTION VS. RSA WEB THREAT DETECTION

Requirement	Traditional Solution	RSA Web Threat Detection
Visibility	Limited view	Holistic view of traffic
Individual Behavior Analysis	Not available	Profile Analyzer Behavior Modeling Crowd -> User Scoring User -> Historical User Scoring
Improved Workflow	Multiple teams Multiple data sources Incomplete data	One tool Holistic view of traffic
End-User Experience	Complex file download	Zero impact until malicious behavior identified
Dynamic Web Sites	Disparate teams unable to update approaches	Dynamic modeling is self-learning and adaptive
Simple Installation	Code install on page Customer software New box in customer path	SPAN Port No customer impact
Time to Production	Days to Weeks	1 hour

Threat Groups

RSA Web Threat Detection protectively identifies threat groups based on similar anomalous behavior and scores. An analyst can give a user-defined name to the threat group and continue to monitor based on the threat group's definition within a top 100 list. This feature also provides control to the fraud analyst to arbitrarily choose a page pattern, monitor that group's activity and take cumulative action.

PROTECT YOUR SITE FROM ONLINE ATTACKS

Catching cybercriminals in the act requires IT departments to look deeper into their web traffic, to examine many more sources of information about web visitors, and to view entire web sessions to determine what website behavior is typical for their website and what is not. Traditional approaches to detecting and preventing fraud don't paint a complete picture of website activity and don't connect the dots between various sources of data about online activity. RSA Web Threat Detection provides this visibility and empowers its customers to stay one more step ahead of criminals looking to exploit their online presence.