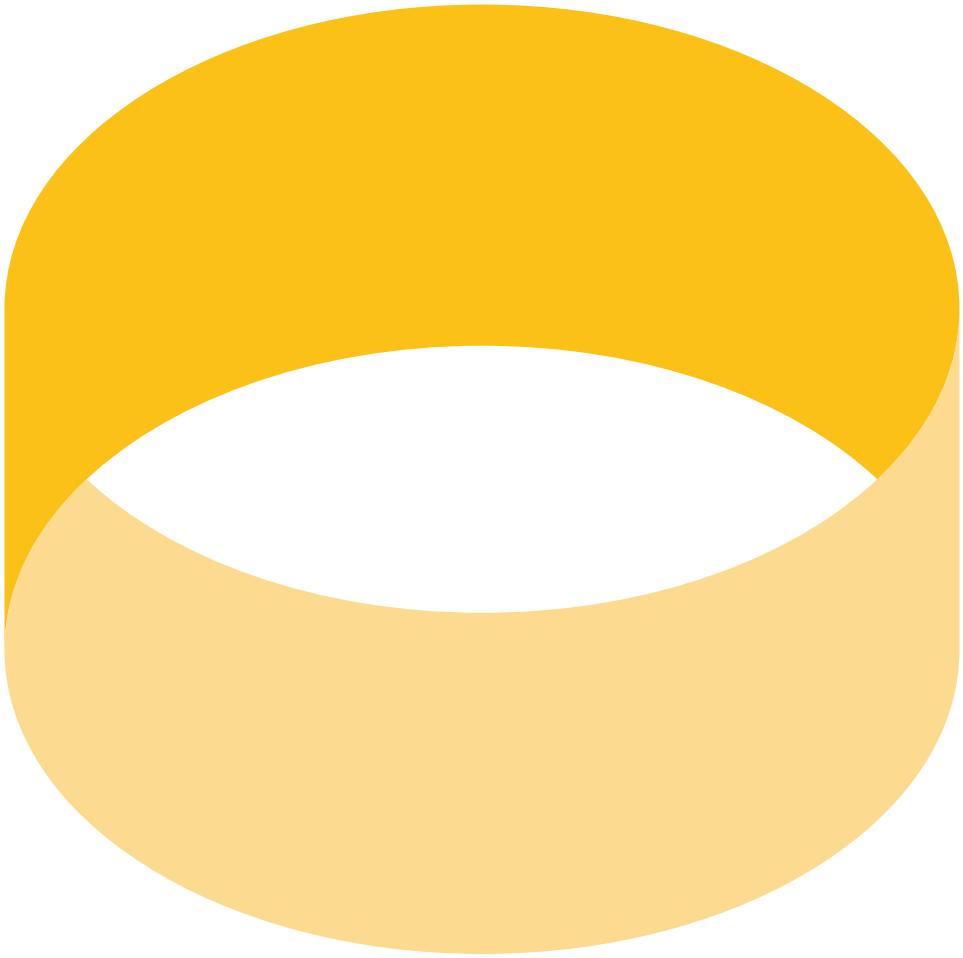


The Need for Threat Risk Levels in Secure Web Gateways

Understanding how a new feature, Threat Risk Levels, helps with uncategorized sites, and how it works with web isolation technologies.

WHITE PAPER

Symantec™

Introduction

Secure web gateways (SWGs) have long been used by organizations to provide secure web access for their employees. In recent years, with the increased threat Level coming from cyber criminals and the increased number of threats coming specifically from the web, SWG technology has become even more important. SWGs have been adapted to the new landscape to detect malicious and suspicious sites in real time, in addition to examining objects downloaded from the web using advanced threat technologies including antimalware, sandboxing, behavioral analysis, and others. Symantec recently announced a new feature, Threat Risk Levels, in our SWG products to help IT administrators deal with the variety of threats coming from the web.

This white paper outlines how the Risk Levels feature works, its underlying technology, and how to use it effectively in SWG policy.

Categorization and Real-Time Rating

Secure web gateways started out utilizing on-device category ratings; on-device databases provided the primary means of identifying malicious and undesirable websites. This allowed for easy policy implementation, such as blocking users from accessing pornography or shopping websites while at work. However, as the number of new websites quickly grew, on-device category databases were no longer able to keep up, resulting in unacceptably high percentages of uncategorized sites. Organizations needed a real-time rating technology to bridge the gap between known and unknown websites.

Real-time rating uses cloud-based artificial intelligence (AI) engines to categorize websites as they are requested by users, significantly reducing the number of sites the SWG is unable to categorize. Sites categorized from the cloud intelligence are included in the on-device database as needed to maximize performance. And the cycle continues...

Symantec's dynamic real-time rating system has evolved to become a highly advanced AI engine in the cloud. It comprises over 300 separate modules that identify a web page's content; categorization occurs in real time, in most instances returning a category in a fraction of a second once the web page contents are retrieved. Some of these modules can identity over 60 languages.

For more information on the Symantec real-time rating system, check out our **white paper on WebPulse**.

The Symantec system currently categorizes each website into as many as four categories. There are 85 categories in all—including 14 generally considered dangerous or potentially risky, including:

1. Malicious Sources / Malnets
2. Phishing
3. Potentially Unwanted Software
4. Malicious Outbound Data / Botnets
5. Spam / Questionable Legality
6. Suspicious
7. Proxy Avoidance
8. Placeholders
9. Dynamic DNS Host
10. Pornography
11. Gore / Extreme
12. Gambling
13. Spam
14. Compromised Sites

Symantec recommends every organization block these 14 "security" categories in their default SWG policy.

Why Uncategorized Sites Exist and Why They Are a Problem

While real-time categorization can classify a majority of new websites, some sites fail to get enough "votes" from the various AI engines to be conclusively classified into a specific category. This may be due to lack of information coming from the website itself; for example, the website may be blank, or it may host only images. In these instances, the real-time rating system returns an "uncategorized" rating (also referred to as "none") rather than a category rating.

While the number of uncategorized sites is relatively small, in a given day's traffic they remain a problem for many IT administrators. Prior to Risk Level ratings and web isolation technologies, there were virtually only two ways to handle uncategorized sites: Allow them or deny them. Each option came with its own problems.

Organizations that wanted tighter security typically blocked access to uncategorized sites. This decision resulted in a greater number of help desk calls from users wanting to gain access to blocked sites. On the other hand, organizations that allowed access to uncategorized sites often found higher infection rates and a greater number of compromised systems, requiring their IT staffs to spend more time remediating affected systems. Neither option worked well; IT administrators needed a better solution.

Symantec introduced Threat Risk Levels to help IT administrators deal with the problems caused by uncategorized sites.

Where Threat Risk Levels Come From

As mentioned earlier, sites are uncategorized because the AI engines do not have enough information—from the website and its associated metadata—to classify the site definitively into at least one category. Yet even without enough information for categorization, there is still quite a bit of information available to the AI engines (generally referred to as our WebPulse system).

A few years ago, Symantec researchers realized that this information, while insufficient to return a category, could be used to help set a Risk Level rating for the website. For example, the history of the website's host IP address could be researched easily; similarly, characteristics of the URL, or the server's behavior, could be analyzed. While any of these factors considered alone wouldn't be enough to rate a site as a higher Risk, they could be combined with other parameters and factors from other AI systems to derive a Risk Level rating.

To provide a Risk Level, Symantec researchers developed three completely new systems to work in conjunction with the existing WebPulse system.

The first new system, the Risk Levels Calculator, is run daily for all entries in our URL database; it's also run whenever new information is added to the URL database. It acts just like its name implies, calculating a Risk Level for every website using a snapshot of the current history available to it, such as the reported ratings and recent traffic. Websites are assigned a Risk Level regardless of category (or lack of category), from 1 to 10, where 1 is the safest and 10 is the riskiest. Organizations can also separately assign a given website a Risk Level of 0 (zero), which is the equivalent of whitelisting.

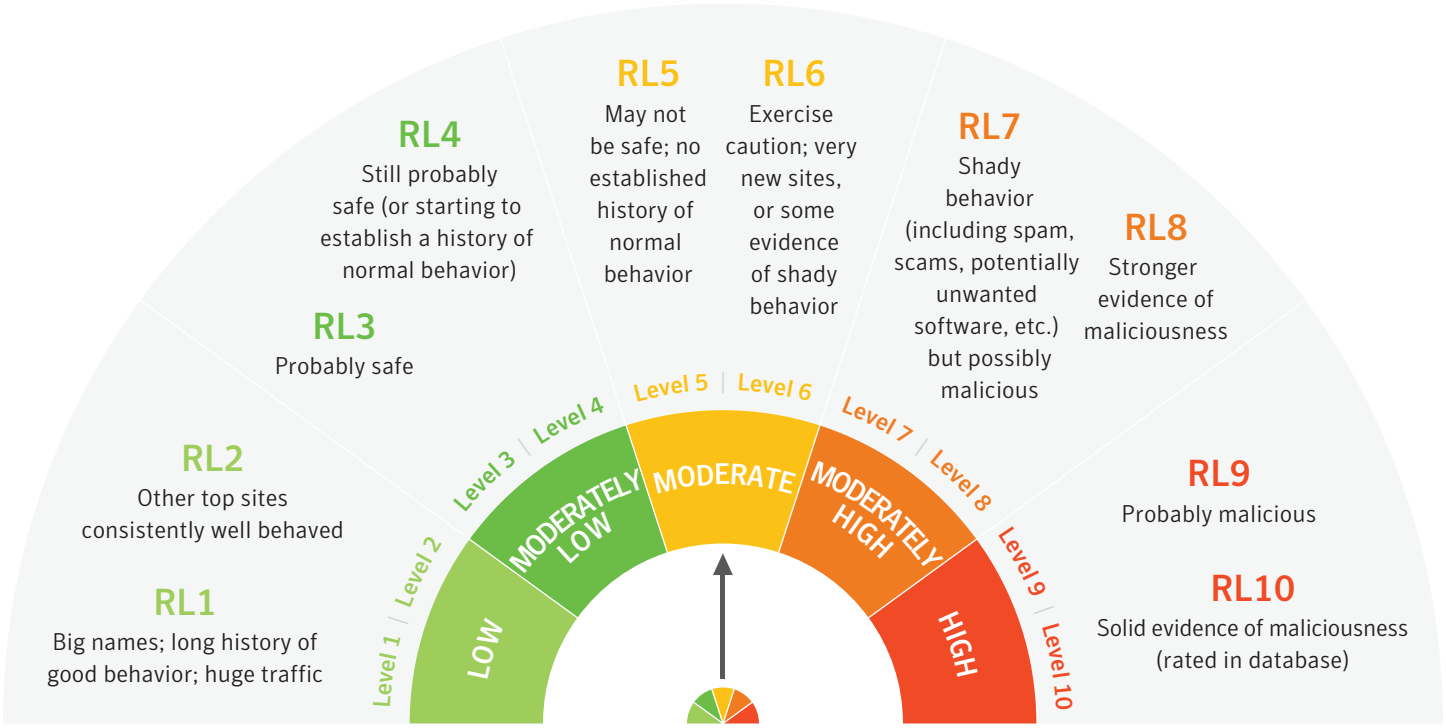
The Risk Levels Calculator works on the URL database, which contains millions of entries. Our data centers actually see hundreds of millions of domains, subdomains, and IP addresses in daily traffic. But a majority of these aren't entered in the database, typically because there are few traffic requests for those sites; they are what we call "one-day wonders," sites that live for only one day or less, or "one-hit wonders," sites asked for only once.

For more information, read our **blog on one-day wonders**.

The second new system, named the "Context Engine," is a massive-scale AI engine used to track, and render opinions about, the hundreds of millions of domains, subdomains, and IP addresses that we didn't store in our URL database. The Context Engine may add sites to the database if its confidence is high enough; it also provides summarized context to the WebPulse system for use in its real-time voting system.

The third new Risk Levels system, WebPulse real-time voting, calculates a Risk Level for all requests not rated by the database. It combines votes from a variety of modules, including the "context" provided by the Context Engine.

Symantec Intelligence Service & Threat Risk Level



RL0
Customer Whitelist

A Brief Explanation of WebPulse and the Context Engine

Symantec started with the problem of having to classify the internet for categories and Risk Levels. Given the massive scale of the internet, accomplishing this required that we combine AI with big data, which we call "Big AI."

As mentioned earlier, we needed to develop Risk Levels for hundreds of millions of subdomains, domains, and IP addresses, but historically we do not have enough information to categorize and place them in our URL database. In our secure web gateway solutions, when a request comes in for categorization and Risk Levels for a website not included in the URL database—and not requested previously be another user—the request gets sent to the WebPulse service.

WebPulse uses a voting system that behaves similarly to a voting system comprising Caucuses and delegates (similar to the voting system used in many primaries in the United States; if you're not familiar with this system, here's a **link to a quick primer**).

When the request goes to WebPulse, the system first checks to see if we already have information on that website stored locally (other organizations may have already asked for the same information, in which case the categorization and Risk Level rating would be cached in the cloud and available immediately). If no information is cached, WebPulse checks with the Context Engine for information on the website. Context Engine has more votes to determine categorization and rating in the overall system than any other part of the rating process.

If there isn't any information on the Context Engine, WebPulse takes that into account as well. Next, the dynamic real-time rating system takes over.

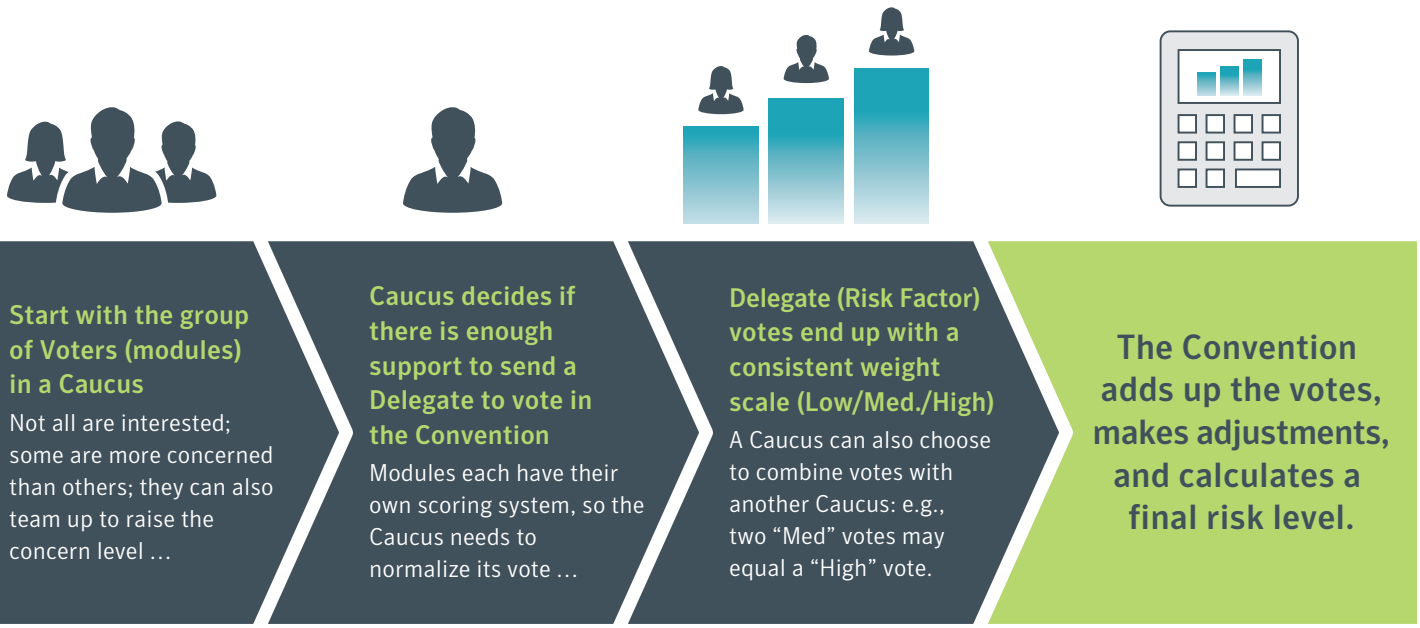
The Context Engine ingests all the data it can find, including the billions of classification requests we get daily. In addition to the WebPulse logs, Context Engine uses a number of feeds from third parties, multiple feeds from other security organizations within Symantec, and other data. In addition to contributing votes on the Risk Level, Context Engine may also have an opinion about a category for a site. The most relevant information to WebPulse is then pushed out from the Context Engine to WebPulse.

WebPulse uses the metadata (or lack thereof) from Context Engine to help it calculate a dynamic Risk Level for every request it receives. The process works something like this:

Any of the WebPulse voting modules ('concerned citizens') can contribute to the Risk Level rating. They are grouped into related areas—for example, modules related to IP address—forming a 'Caucus'. Approximately 20 Caucuses are allowed to vote in the current dynamic voting system.

WebPulse Dynamic Voting System

Risk factor data analyzed by "voters" grouped into "caucuses"



Each Caucus can send a 'vote' to help set the final Risk Level rating assigned to the website. The Caucus individually decides if there is enough support to send a 'delegate' (vote) to the 'Convention'.

Each of the modules in a Caucus may have its own scoring system, so the Caucus needs to normalize its vote. Delegate (risk factor) votes end up with a consistent weight scale (low/medium/high). A Caucus can also choose to combine votes with another Caucus; for example, two medium votes may equal a high vote.

Each Caucus vote goes into the final part of the system, which, to maintain the political theme, we call the Convention. During the Convention, WebPulse analyzes and scores the top four votes—a result that takes into account which Caucuses voted, and how strong the vote was—from the Caucuses. The Convention adds up the votes, makes adjustments, and calculates the Risk Level.

These Risk Levels are calculated with every request, so any significant new data can change the Risk Level. Finally, WebPulse shares the overall Risk Level and one or more content categories, if any could be determined, with the requesting secure web gateway system.

A sample of the Caucuses and modules involved in the dynamic voting system is shown in the below 'Stages of Analysis' diagram.

Stages of Analysis

STAGE 0 (Database)	**Database has Category and Risk Level. Later stages always produce a Risk Level (and a Category if possible).**	**Database is fed by Research, Background Checks (see Stage 5), and Feeds (including Customers) ...**
STAGE 1 (Request)	**Everything starts with a Request (we have the full URL and the HTTP headers to work with).**	**Includes:** • Shady Traffic Group (Traffic Cops, etc.) • New/Untrusted Group • Shady Name Group (Weirdo, Ghostbuster, Shady TLD, etc.) • Shady Hosting Group (DynDNS, FreeHost, etc.)
STAGE 2 (DNS)	**Now, we have the Resolved IP Address to work with ...**	**Includes:** • Server DNA Group (Malnet Tracker, etc.) • Shady Neighborhood Group (IP, Range, etc.)
STAGE 3 (Response)	**Look at the Response from the destination server (including the Content it returned).**	**Includes:** • Shady Response Group • Network Error Group • Lie Detector Group • Shady Content Group (EXE, JAR/ZIP, PDF, JScript, ...)
STAGE 4 (Post-Analysis)	**Take a second look (or a deeper look) at everything we got to this point.**	**Includes:** • Phishing Detectors • Metarule modules • Yara Rule modules **This is also where the "Caucus and Convention" voting system fires.**
STAGE 5 (Background)	**We want to gather, collate, and analyze as much data about the site and its behavior as possible ...**	**Includes:** • Malnet Tracker Group • Site Profile Group (via Context Engine) • IP Profile Group (via Context Engine) • Hacked Site Group (via Context Engine) • Chatter Group (via DB) • DeBRa Group (Deep Background Raters)

Web Browser Isolation Primer

The Risk Levels feature isn't the only new one available to help with uncategorized sites. Symantec recently announced the acquisition of a web isolation company, Fireglass. Web isolation (also known as "remote browsing") is a fundamentally new way of handling potentially dangerous websites.

Instead of executing the contents of a requested web page directly in the user's browser, web isolation technology executes the web page using a remote, cloud-based virtual browser, and delivers safe rendering information of the page to the user's browser. By executing the web content remotely, the web isolation service prevents any malicious code from reaching and executing on the user's system.

Find additional, detailed information on Symantec Web Isolation: **www.symantec.com/products/web-isolation**.

Using Risk Levels and Web Isolation for Categorized and Uncategorized Sites

Used together, Risk Levels and Web Isolation are powerful tools for handling both categorized and uncategorized websites.

First, let's look at the uncategorized problem we described earlier.

For sites not in the database, WebPulse will be consulted and a Risk Level from 1 to 10 will be returned. We generally consider a site with a rating from 1 to 3 to be safe, and we recommend an organization's policy allow users to freely browse that site. Organizations with extremely sensitive security requirements can set up a policy to send this traffic to isolation, so the user is further protected from any potential risk.

With a rating from 4 to 6, there is some suspicion regarding the safety of the website. Here the recommended course of action is to put in a coaching page recommending that users do not go to the page—the coaching page would typically offer two options: to leave and to proceed to the site—and notifying them that if they go to the page, their activity will be monitored. In this instance, we recommend the use of web isolation technology regardless of the organization's security stance.

For the remaining Risk Levels, 7 to 10, we recommend complete blocking of these sites, as they have indicators of high Risk. For organizations that absolutely need to provide access to these sites, web isolation is warranted, as is blocking any downloads.

Organizations that want a higher security profile, and are moving to Risk Levels for greater security, should consider setting a block Level starting at Risk Level 6. Highly sensitive organizations with skilled security teams familiar with WebPulse should consider blocking starting with Risk Level 5. This recommendation is also true for organizations that have historically blocked uncategorized websites.

Now we'll look at categorized sites.

Certain categories are inherently risky; for example, it's been well documented that the 'Pornography' category has a high percentage of malicious sites. The Risk Levels feature gives an added dimension to any website regardless of categorization, and offer an additional tool in deciding how to proceed with data-sensitive web categories such as 'Healthcare' and 'Finance.'

In the cases of 'Healthcare' and 'Finance' specifically, many organizations have recently needed to protect users' privacy regarding their records. This need competes directly with the emerging requirement to inspect encrypted web traffic, which cyber criminals are increasingly using to distribute malware and to collect personally identifiable information. Risk Levels enable the IT administrator to allow safe sites (rated 1 to 3) to be accessed without SSL inspection—thereby protecting user privacy—while decrypting riskier sites for additional inspection, including advanced threat detection and data leakage protection.

Other categories that benefit from web isolation technology are webmail and online file storage. For example, Gmail and Dropbox are rated as Risk Level 1 sites, meaning the site themselves are safe; that is, extremely unlikely to be hacked—but these sites could be hosting unsafe files. Because unknown content is stored on these sites, these categories are ideal for use with web isolation, which allows for remote viewing of documents without the need to download them to the local machine.

Another category presenting an interesting use case is web ads/analytics. These sites tend to have a wide array of Risk Levels. It makes sense for organizations concerned about "malvertising" attacks to block these sites more aggressively—starting at Risk Level 4 or 5—since malvertising is more likely to involve these slightly higher Risk Level ad sites.

Conclusion

Risk Levels help solve problems faced by many organizations using content filtering technologies on their user's web access. Risk Levels give organizations more flexibility in handling user's requests to visit new and unknown websites—without the typical problems of completely blocking access to uncategorized sites, or the increase in compromised systems that results from completely allowing access to uncategorized sites. Risk Levels also provide a way to effectively use new technology, like Symantec web isolation (formerly Fireglass). Risk Levels are found in Symantec's Intelligence Services offering, and secure web gateway products including ProxySG, Advanced Secure Gateway (ASG), SG-Virtual Appliance, and Web Security Service.

Symantec

350 Ellis St., Mountain View, CA 94043 USA | +1 (650) 527 8000 | 1 (800) 721 3934 | www.symantec.com

20B277874_WP_SWG_RiskLevels_EN