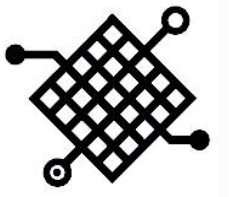
ChainLink Halo

Protecting your investment on the Web



Web Site Server Monitoring

Summary

To maintain an effective presence on the web, it is essential to collect useful web traffic data and metrics in order to maintain the highest quality web experience for users. The data collected should provide insight into how well a web site is performing and what the users were doing while on the web site. To provide a true indication of the user experience, it is necessary to record all aspects of the client server transaction. This paper presents what web site administrators are faced with when attempting to handle these challenges and how the ChainLink Halo meets these challenging requirements.

Challenge

When web site administrators are faced with the task of monitoring a medium to large-scale web presence, they quite often rely on multiple independent tools to log traffic, analyze user activity and determine the overall health of the web site.

Some of the web server traffic can be observed through analyzing web server log files. The web server log files will normally identify the hit time, client IP address, requested uniform resource locator (URL), server response code, response size, referring URL, client cookie and browser. Many commercial log analyzers exist that the administrator can use to collate the log files and report on general traffic patterns. Since log files can grow quickly to be several hundred megabytes each, analyzing web traffic through log file analysis becomes difficult for busy web sites.

Also, as the popularity of the web service grows, the ability to pinpoint individual statistics becomes more difficult and a relational database is required to enable the administrator to filter out unwanted statistics and create custom reports. The problem is that adding database logging within a web server causes additional load on the web server as well as the network both of which can impact the quality of the user experience.

When a problem occurs on a web site, the administrator needs to reconstruct as soon as possible what happened so that the appropriate corrective action can be taken. Log file analysis can provide some assistance to alert an administrator to problems such as bad links and server errors, but they often leave an incomplete story.

Another tool that administrator's use for trouble shooting network problems are packet LAN analyzers. The LAN analyzer picks up any traffic that exists on the network and saves it to a file. Again, the more busy a web site, the faster the file can grow, and without very narrow filters in place, they very quickly become too big to be useful. Additionally, even if the problem transaction can be identified, making sense of the captured packets can be tiresome.

For monitoring server availability and performance, web administrators use pre-defined link testers to sample parts of the web site. Metrics for response times are collected and errors are logged. If an error occurs, the request and response headers as well as the error page can be saved and reviewed later. There are two problems with this approach. One is that if the sample is

small, it does not accurately represent the site. Secondly, if the sample is large, it adds unnecessary traffic to the network as well as add ional load on the web server.

ChainLink Halo

The ChainLink Halo is a passive network appliance server, which sits quietly on any ethernet segment and collects HTTP traffic for use in the marketing and administration of your collective web hosting services. ChainLink Halo provides advanced monitoring and data collection tools to facilitate the monitoring of multiple web sites simultaneously without impacting traffic or web site performance.

The ChainLink Halo contains three core components that combine features of the most common tools that administrators use to monitor their web servers. The three components are the ChainLink WebWatcher, ChainLink WebLogger, and ChainLink WebAnalyzer. The components are packaged together in a single ChainLink Halo or can be distributed as separate agents across multiple application servers. All three components are managed through a single web application so that they can be accessed remotely.

The ChainLink WebWatcher is a packet capture component that splits apart HTTP specific data into request and response header and data streams. It records important metrics about each part of the web transaction and will save any or all pieces of the web transaction to either file or directly to the ChainLink WebLogger. All servers that are on the same LAN segment can be monitored simultaneously. The ChainLink WebWatcher aggregates all web server traffic into one central repository. Web traffic log files can be effectively turned off on all web servers so that the servers are free to do what they do best, server web pages.

The ChainLink WebLogger is the central repository for all web traffic data. Multiple devices can log data to the ChainLink WebLogger including remote WebWatchers, ChainLink Halo data files, web servers or web server log files. Periodic cleansing and compression as well as smart caching of identical objects minimize the storage requirements for the ChainLink WebLogger.

The ChainLink WebAnalyzer catalogs the captured web transactions into logical groupings so that the web transactions can be easily located. Each catalog has user-defined indexes and rules for filtering content as well as how long the content is to live with the catalog.

For example, one catalog could be defined to capture only web server errors and the entire data transaction for both request and response would be saved and stored for 24 hours. An administrator can periodically check (or be alerted via email) on the server errors and review the entire transaction to determine the cause of the problem.

Another catalog could be defined to capture only web server request information to be used for marketing purposes, which is needed for 1 year. Since the data objects are not needed for marketing purposes, they are ignored for that catalog. Within the header request, a cookie is set which identifies a user uniquely. That user id is parsed from the cookie and indexed as part of the catalog logged data so that all user activity can be uniquely identified. Any external log analyzer software can be used with the ChainLink WebAnalyzer, so for instance, WebTrends could be used to generate reports from this catalog for any period filtered by user or any other filter.

Other catalogs can be set up to catalog unwanted or unauthorized intruder access activity, special promotion areas, individual web sites, etc.

When all three components are deployed together on a single Web Halo, no additional bandwidth is incurred for monitoring web traffic.

Although many predefined filters have been included in the ChainLink WebAnalyzer, additional ones can be easily created and integrated. The ChainLink WebLogger can be accessed through JDBC/ODBC drivers so that custom reporting tools can take advantage of the data gathered.

Conclusion

Web site monitoring while essential, can be ineffective without the proper tools. As web sites become more and more popular, the challenge of monitoring becomes even more difficult. The ChainLink Halo network monitoring appliance provides a comprehensive set of tools which simplify an administrators ability to monitor the traffic on medium to large scale web sites. Additionally, the web traffic measurements gathered with the ChainLink Halo provide added value statistics for marketing purposes and the ability to generate micro and macro level reports with user defined filters.

ChainLink Networking Solutions, Inc.

11012 Burywood Lane
Reston, VA 20194
Phone (703) 430-5752
Email: info@chainlink.com