BPA Interactive Audit Methodology

This document describes the methodology through which BPA audits Web site traffic. The resulting audit report will include usage data from the analyzed Web site traffic, and may also include demographic data collected from users during the registration process.

BPA employs a census methodology that requires the analysis of all recorded traffic for the entire specified audit period. Log files containing the recorded traffic are accepted in various formats. BPA audit staff conduct all analyses.

I. Audit Process

Data Transfer:
Logs are uploaded to BPA via an FTP account, in a compressed format.

Processing General Traffic:
An auditor processes the transferred logs using a series of custom-written programs. These programs:

- Filter traffic by excluding requests for supporting files, such as images, audio, video, style sheets, most media that require plug-ins (PDF files are among the exceptions), and program files that run in the background.
- Exclude automated traffic by filtering spiders/robots (e.g., Architect Spider).
- Requests are excluded based on the method used to request content (e.g., HEAD).
- Requests are excluded based on the status of completion. At this time, only 200 and 304 HTTP status codes are accepted. The remaining codes are incomplete loading codes that are excluded.
- The filter process produces a report that verifies that BPA has received data for the entire audit period and can provide evidence of additional possible spider activity, or express the possibility of when servers were not operational. If traffic data is found to be incomplete, an explanatory comment will appear on the audit report.

Processing Ads:
If the audit is reporting ads that were requested, these are processed separately from general traffic usage, but follow the same rules (with the exception of certain file types).

Duplication Checks:
During processing, the traffic is analyzed to determine if there is any duplication. If duplication is found, the affected lines are excluded from the analysis.

Data Storage for Analysis:
A database is created and the filtered traffic is imported in preparation for analysis. Once data is imported, additional filters may be applied and the imported traffic is analyzed. The analysis produces a report showing different points of view of all the activity during the specified audit period. These views provide data for the audit report, as well as information used by auditors to find any additional traffic that should be exluded from the final analysis.

II. Additional Integrity Procedures

Seeding:

Auditors visit sites for the purpose of "seeding" and record the seed activity. When log files are received, the auditor extracts the relevant records from the log files. A comparison is done to determine if:

- the site's server is accurately recording the time and day.
- the site's server(s) are accurately recording activity for the site.

If discrepancies are found, they are investigated and reductions could occur. If the activity is correct, BPA concludes that all information from users has been logged accurately.

BPA reserves the right to refuse to conduct an audit if we find that the log file(s) have been tampered with.

During analysis, additional filters, such as host addresses and/or pages, may be applied if the auditor has determined that the traffic should be excluded after visiting the site. Standard filters include the site's own host addresses, BPA's host addresses, robot/search host addresses, Java scripts, audio files and any framed pages or redirects that are found during seeding or analysis.

- **Framed Pages**
  When Web sites are created, some designers employ framed pages. This means that when you request a single page, the server delivers up a group of frames as one page. The server then records each one of the frames as an individual request in the log file. These frames are filtered during analysis so that over-counting of activity does not occur.
- **Redirects**
  Advertising banners on Web sites usually redirect the user to the advertiser's Web site. These redirects are excluded from the analysis so as not to count access to other sites or double-count activity within a site.
- **Robots/Search Engines**
  Search engines access sites in order to provide Web site links each time a search is requested. Access to the site is recorded by the hosting site, but this activity is not deemed a legitimate request for audit purposes, since a human being did not view the requested page (rather, a computer hit the page in order to collect site content). The search engine returns links to related sites based on search criteria entered by a user. When a user accesses a site by selecting a link returned by the search engine, that request is also recorded, but is counted as a legitimate request.
- **Java Scripts**
  Java scripts animate graphics or allow for scrolling text on Web site. These files (e.g., .class, .map, .ncb and .jar) are excluded from the analysis because they are not legitimate page requests, but rather an application that runs within the page.
- **Audio Files**
  Audio files are filtered out because they do not constitute a legitimate page request. They are either downloaded off of a page or "stream" when a page is requested. Streaming of an audio file means that the sound automatically loads with the page. Audio (and video) may be included in the audit report, but are identified separately from page requests and visits.
- **Downloaded Files**
  When a user downloads a file from a Web site, a request is recorded. There is usually no way to determine if the downloaded file is ever viewed. These requests are excluded from analysis, as they are not in an open state and are not interactive with the Web site.

III. Audit Data

The audit results may contain the following information: page requests and/or visits analyzed daily, hourly and by day of week. A listing of the most-accessed pages and the most frequent sources of traffic may be included as an option. These results are then tested in the following manner:

1. Daily, hourly and day-of-the-week results are examined to determine any spikes in activity. Unusual activity changes are investigated to resolution.

2. The top pages requested are viewed via Web browser to ensure that they are legitimate pages.

3. The most frequent sources of traffic are identified to determine that they are not associated with the site in any way, and that they did not visit the site using an automated procedure.

These procedures allow BPA to verify that only legitimate pages are being counted in the activity; that a site owner is not increasing their traffic numbers by visiting their own site; and that all activity is accounted for throughout each day of the audit period.

IV. Registered Database

Audits of a registration database are conducted at the request of the site, for an additional fee. Complete demographics of the registered-user database are then included in the audit report. Interested parties should contact BPA to apply for a database audit or for inquiries.