



3 ways of measuring site traffic advantages & disadvantages

Tom Breur
September 2005

Intro

Measuring site traffic is vitally important to have numerical support for allocating web design budget. How do you prioritize if you don't know *how many* and *what kind* of visitors are accessing certain pages? Further, in-depth analysis can increase site effectiveness and segment user types. For these purposes one needs to capture site usage data. This can be done, fundamentally, in one of 3 ways:

- Server based logs
- User based logs
- ASP based logs

All 3 methods have advantages and disadvantages to offer. In this whitepaper the author will allude to the benefits and drawbacks of all 3 methods. Also, some technical clarification will be given to explain *why* each method has its own merits.

Server based logs

The physical server on which a website is hosted produces a detailed account of each page request made. These data are generated as a 'byproduct' of the server's operating process. Because web usage analysis was never envisioned when these systems were designed, the nature of these data makes it rather cumbersome to transform these data into a form that is usable for analysis.

To make such data useful for analysis, a mapping of the 'raw' server data to an appropriate data model requires substantial domain knowledge. In principle the data are very 'rich', but it is labor intensive to extract the full value embedded. Also, the server needs to be configured appropriately in order to generate all the data fields that may be of interest. Reconfiguring is not very difficult, but does require IT involvement.

Especially when the site hosting is outsourced, such reconfiguring may well interfere with the host's needs. One machine can host several other sites. Raw server data have a voluminous layout. For a large site this runs in the Mega- up to the Giga bytes per day. These data need to be stored and transferred to the analytic users. Copies, backup and maintenance of such data needs to be negotiated with the host, often a third party resource.

In conclusion, despite the fact that these data are basically generated for free as a byproduct of website operation, getting timely physical access to the right data, in the right format can be challenging.

User based logs

With user based logs, the web server drops a very small program on the user's computer for the sole purpose of tracking clickstream data on the client's machine. This technology has many advantages that are mostly derived from the fact that such monitoring programs were developed for the explicit purpose of capturing clickstream data. This allows very 'rich' data in the most useful format. Also, because measurement is performed on the client machine, you get the ultimate client centered data. In contrast, measurement with server- or ASP-based mechanisms is inevitably subject to slight misinterpretations of the actual time lags the user is experiencing. This is caused by the fluctuating speed of the connection between client and either server or ASP.

However, in recent times, some website owners have developed such applications for rather intrusive and often unpleasant purposes. This technology has come to be known colloquially as spyware or adware. Most consumers are (rightfully so) weary of having such applications installed on their computers. Consumer distrust basically precludes this technology from being used, unless the websurfer has given consent in advance. Such willingly and purposely engaging in tracking of surf behavior is the case for consumer panels (like for instance Nielsen NetRatings) where people voluntarily participate in research.

In conclusion, this method is useful and powerful, but due to lack of consumer trust is not applicable anymore in most cases. Only after consumers have provided explicit consent in advance is this method a viable option. Tracking surf behavior of employees on an intranet may be done in this way, for example.

ASP based logs

With ASP based logs, a dedicated, separate server is put in place to track visitors' surf behavior on designated sites. To this end a so-called 1-pixel gif file is placed on each page of the site. This is non-intrusive and invisible to the user.

The ASP server is set up for the sole purpose of tracking surf behavior (clickstream data). Therefore, server configuration and data formatting can be optimized to this end. Naturally, this allows for capturing all the relevant data, and ignoring any of the 'fills' that are useless for analysis purposes. This results in much less voluminous storage and optimal accessibility of data. The data are modeled in a format that is easily amenable for analysis.

Conclusion: with only the requirement of placing a tracking pixel on each site page, this method requires the least effort from the site hosts. Data storage and maintenance are facilitated due to the dedicated nature of the ASP server. Data are always quickly available, and pre-processing can be completely automated. The ASP server can be maintained in-house or outsourced. Most professionals in the field consider this the method of choice.