WAVECREST
C O M P U T I N G

Hits vs. Visits: Internet Monitoring the Right Way

Wavecrest Computing

904 East New Haven Avenue
Melbourne, FL 32901
Toll-free: 877-442-9346
Voice: 321-953-5351

info@wavecrest.net
www.wavecrest.net

Abstract

In the rapidly burgeoning "e-world", it is becoming increasingly important for enterprise managers to know exactly how their network resources are being used. Such awareness is essential if they are to configure, plan and control those resources in ways that optimize performance and avoid legal and personnel problems. In this environment, the most significant component is the World Wide Web and the activity that takes place on it. To keep track of Web activity, it is important that management receive comprehensive, detailed reports that accurately and precisely reflect the volume and character of that activity. Unfortunately, such precision and accuracy are hampered by confusion over terminology. In particular, the terms "hits" and "visits" appear to be widely misinterpreted or confused, a situation that can lead to serious managerial problems.

This paper discusses these terms, elaborates on their meaning and usage, and offers some definitions. In the author's view, if these definitions could be brought into general use, they could help alleviate some of these problems. The author also points out the importance of report designers making a clear distinction between visits and hits on their products.

To keep track of Web activity, it is important management receive comprehensive, detailed reports that accurately and precisely reflect the volume and character of that activity.

Introduction. In the wonderful, wacky world of the World Wide Web, everyone loves to talk about "hits" and "visits." The comments and questions go on and on. "Who visited what sites?" "How many times was Amazon.com hit yesterday?" "Is Charley really spending time on sex sites?" Some times these questions are trivial. At other times, though, they can have important implications. Just think about businesses, schools or other institutions that grant Internet access to large numbers of their personnel (we refer to such personnel herein as Web site "visitors"). In such organizations, which we refer to collectively as "enterprises", the question of how such access is being used can have significant implications. These range from cost issues to profit, morale, productivity, social and even legal concerns. Think also about other enterprises, located on the other side of the "firewall" so to speak, that rely on their Web pages to attract customers, suppliers or other enterprise-related visitors.

Within both types of enterprises, there is much discussion about Web activity. There is also much *reporting* of such activity. A few years ago, the terms "visits" and "hits" came into use to facilitate this communication process. But do they? And what do these terms really mean in the context of the World Wide Web? And because of the huge investments in Web activity and resources, and their importance to the organizations that rely on them, shouldn't these terms be universally understood, used with precision, and interpreted in the same way by different people?

Unfortunately, that doesn't happen all the time. Many people frequently confuse the two terms and often use them interchangeably. This can be very misleading, particularly when the speaker or writer has *one* meaning

Wavecrest Computing

904 East New Haven Avenue
Melbourne, FL 32901
Toll-free: 877-442-9346
Voice: 321-953-5351

info@wavecrest.net
www.wavecrest.net

in mind, and the listener or reader has *another*. Such confusion can result in the listener or reader drawing erroneous conclusions. And erroneous conclusions can lead to poor decision-making and ill-founded actions.

OK, so what *do* these terms mean, and what *are* the definitions? Well, that's the purpose of this paper, i.e., to explain the terms as they are generally and technically used and to attempt to provide clear, useful definitions. Before doing so though, it may be helpful to provide some background on Web sites (and Web pages) with emphasis on the way in which computer users reach them and obtain information from them.

Web Sites. A Web site is a computerized, screen-viewable electronic "publication" that consists of one or more "pages". The first page is often referred to as the "home page". Once at the home page, the "visitor" can access subordinate or related pages by clicking on "hyperlinks." If a visitor knows the linked page's address, he or she can by-pass the home page and go directly to it. Web pages contain "content", i.e., subject matter information, the visitor is presumably seeking. Additionally, Web sites often display various computerized banners, ads, and multimedia presentations. Typically, these items are received automatically from *other* ("third party") sites when the visited site is accessed. These items usually constitute some type of advertisement, conceptually similar to newspaper or magazine ads. Their appearance on the screen occurs automatically regardless of whether or not the visitor cares to see these items. Lastly, Web sites also frequently make "downloads" available to visitors. Downloads can be computer programs, compressed files, digitized movie films, sound clips, etc.

Web Site Creation. Web sites are created by people called Webmasters or Web-site "authors". These creative individuals imbed the content in the pages, make downloads available via "click buttons" and arrange for the importation of banners, ads, etc. Once they finish the design, they arrange to have their sites placed on Internet (or Intranet) "servers". Servers are simply large computers that "host" many sites. Servers make these sites available to computer users who have network access.

Accessing Web Sites. Computer users can access and view Web sites and pages via "browsers" installed on their desktop computers. A browser is simply a computer program that can obtain, decode and display Web site information on demand. It springs into action when the user transmits the desired site's address (e.g., "http://www.amazon.com") to the network. You might think of this whole process as "going to the site for a visit."

> **Note:** IT specialists often refer to browsers as "clients," "client browsers" or "Web browsers." They also frequently refer to site addresses as "URLs." Don't let the lingo throw you. The concepts are simple.

Confusion Factors. Don't worry, I haven't forgotten the objective of this paper. We'll soon get around to defining "visits" and "hits." But first, let's look at some of the factors that confuse the issue.

When the user "goes to" the site, he obviously intends to engage in a *visit*. (Let's think of that action as a single, "clickable", *manual* event.) But as soon as he or she clicks the mouse (or hits "ENTER"), a number of banners, ads, multimedia presentations and even download links may automatically "go to" the same Web site address and then be relayed to the visiting user's screen. Because each of these "arrivals" has its own URL, all of them — those that were triggered manually and those that were triggered automatically — are individually recorded for statistical purposes at the server location. Such automated records, known as log files, make no distinction between manual and automatic events. *(Make a note to remember the two preceding sentences. We'll be returning to this subject later.)* Furthermore, in a sense, each event is registered twice — once with respect to the Web site itself, and once with respect to the "visitor."

Viewed this way, you can see that one click by one user can result in the logging of many events. Often referred to collectively as "activity", this swarm of events receives a great deal of attention —from many quarters. Such attention comes from the visiting users' camps as well as the Web site camps.

It's time to return to the terminology problem that we started out to resolve. That is, does all of this manual and automatic Web activity constitute hits, visits, or both? Well, as former President Clinton once said, "It all depends on your definition." So we'll just have to create some definitions. Hang on though. We're still not quite ready. Here's some more background.

Part of the confusion over "visits" and "hits" stems from the fact that, without counting the individual users, there are three groups of people who are vitally concerned with Web site activity. Unfortunately the three groups come from different disciplines, have different agendas, and sometimes speak different "languages" when it comes to Internet usage. Let's take a brief look at these three groups.

1. The first group, which is "visitor-oriented," is concerned only with manual events. Typically, people in this group are line managers, HR personnel, and members of the legal staff. They tend to think of Web activity as content-driven "visits" made by human beings. Their concern is with their computer users' actions. Which sites are they visiting? How many visits did they make? Are the visits appropriate? Web activity statistics are important to this group, and members generally assume that one event constitutes one visit and is recorded and reported as such.

2. The second group — located "on the other side of the firewall" so to speak — is concerned with the design, performance and attractiveness of the Web site itself. People in this group tend to be business development or Webmaster personnel. Their perspective is somewhat different from that of the first group. Primarily for business and cost reasons, they tend to be concerned with *both* types of events, manual *and* automatic. Some of these individuals may tend to blur the distinction between manual and automatic events and refer to all of them as "hits."

3. The third group consists typically of Information Technology (IT) and financial personnel. Like the people in the second group, these folks are concerned with *all* events, both manual and automatic, but from yet *another* perspective. They know that network activity of any kind can impact system capacity considerations ("bandwidth consumption", in geek-speak). And system capacity considerations can have significant cost and technical implications. Consequently, people in this group generally see things this way: (a) all events constitute "activity"; (b) all activity should be counted and recorded; and (c) site content is not particularly relevant to their concerns. These individuals tend to think of every event as a "hit."

Herein lies the rub. We now have a "Tower of Babel", populated with three different types of people, with different agendas, who often use different "languages." They all need to communicate with each other, but they see network access and Web site activity from totally different perspectives. Unfortunately, their use of different "languages" can lead to serious misinterpretations. For example, when a person from the first group (user perspective) is talking to someone from the second or third group (Web site or IT perspective), and the first person is referring to "visits" while the listener is thinking "hits", communication is seriously distorted. The reverse is also true. A person from the third group may be talking about "hits", and his audience (from the first or second group) thinks the speaker is talking about visits.

Here's another complication. Remember log files? They're the automated records of computer-user and Web site activity that we spoke about earlier. If someone, e.g., an HR person investigating allegations of excessive misuse of network resources interprets all records as "visits", he or she could be seriously misled. Ill-founded conclusions based on such interpretation could result in misguided action, e.g., reprimand or termination of an employee. In such a case, the mistake would stem from the fact that one manual event can trigger many automatic events (banners, ads, multimedia presentations, etc.).

Similarly, a Webmaster or business development person can be seriously misled if he or she equates all logfile records with "visits". How many times do we hear, "Such-and-such a site was "hit" 200,000 times

yesterday!," conveying the impression that 200,000 people visited the site. If a Webmaster or business development person makes this incorrect interpretation, and relies on that interpretation as the basis for some decision or action, disaster could result.

Well, I think we've finally come to the end of the confusion factors! It's time to cut through all this verbal fog and come up with some simple definitions. Then all we have to do is hope that people will use them.

Definitions

Introduction. I can see it now. You're probably thinking: "How presumptuous! Imagine! Writing definitions for the whole world to use!" Well, be advised that I first tried to find some suitable standard definitions already in existence, particularly for the terms "hit" and "visit." After all, if I could, it would save us all a lot of trouble. Well, while I didn't research every possible source, I did try some typical computer dictionaries, and I wasn't too successful. The best I could find was a definition of "hit" which read: "A successful match or search of a database." That's OK, but not very useful for our purposes. And I couldn't find *any* definition of "visit" in the computer dictionaries. So, I decided to plunge ahead "where angels fear to tread" and create my own. Read on.

Hit. (Noun) Any type of viewable or usable data transmission that is triggered by a visit to a Web site and is received by the computer user's browser. Such transmissions can be in the form of a file, message, object, graphic, link, banner, ad, or push item. For discussion and reporting purposes, these items are classified as "visits", "downloads", or "unsolicited objects."

Visit. (Noun) A "visit" is a "virtual trip" to a Web site. It can be thought of as a manual "click-action" intended to enable an individual user ("visitor") to view or download information contained in the site. The initial click-action constitutes one "visit." If the visitor clicks on other locations within the site (links, downloads, etc.), those subsequent clicks are *also* considered to be visits. The important point here is that one "trip" to a Web site can conceivably constitute one or several visits. (It can also lead to or trigger some other events, as discussed in the next two paragraphs.)

Download. (Noun) A download is one of several types of items that a computer user can "import" into his system from a distant computer or server. A download can be a computer program, graphic object, compressed file, digitized photograph, animation clip, movie file, etc. The user initiates the download action via a click on the Web site that he or she is visiting. Each download is considered a "hit."

The term "download" is also used as a verb. When it is, it refers to the act of transmitting information from one computer or server to a (usually) smaller, individual user's system. Note that the term "upload" means just the reverse.

Unsolicited Object. (Noun) An "Unsolicited Object" is one of several types of information items that appear on a Web page but are not part of the *main* content of the site. Typical unsolicited objects are banners, ads, multimedia presentations (animation, video, etc.), and "push" items, e.g., Microsoft channels, Pointcast, and Backweb technology. Although they appear "automatically" on a Web page, they have their own URLs. Third parties typically pay the Web site owner to incorporate these objects into the site, usually for advertising purposes . just like ads in magazines or newspapers. When a user initiates a Web-site visit, unsolicited objects are routed to his or her screen by way of the Web site server (where they are integrated with the basic content). As they pass through the server, each objects is treated as a separate URL, log-filed and counted as a "hit" (on the site). These same log files will treat these events in a way that counts them as "hits" on the visitor's *browser*.

Post Script

So, aside from establishing some working definitions, what are the *implications* of all this discussion? Probably the most significant is this. Reports and audits that characterize or depict Web activity must clearly distinguish between *true* visits and total activity, i.e., "hits." To be sure this is the case, people who design Web activity report formats should be very precise in labeling the data, and they should include appropriate explanatory notes to preclude misinterpretation.

Web activity report formats should be very precise in labeling the data, and they should include appropriate explanatory notes to preclude misinterpretation.