Managing Internet Usage with Reliable Metrics

Wavecrest Computing
904 East New Haven Avenue
Melbourne, FL 32901
Toll-free: 877-442-9346
Voice: 321-953-5351
Fax: 321-953-5350

www.wavecrest.net

Preface/Abstract

This paper is intended to help IT personnel, business managers and HR professionals understand, select and implement reliable Web-access reporting systems to augment their Internet usage policies. When carefully chosen, implemented and integrated, Internet usage policies and their companion reporting systems can help businesses in several ways. Most significantly, they can help maximize work force productivity, minimize resource costs, preclude security problems and avoid legal liabilities.

Companies that do not monitor employees' surfing habits make themselves vulnerable to legal liabilities, probable bandwidth abuse and employee productivity gaps.

> Companies that do not monitor employees' surfing habits make themselves vulnerable to legal liabilities, probable bandwidth abuse and employee productivity gaps. — *The Aberdeen Group*

In discussing these issues, the author emphasizes several key points:

- To be truly useful, usage policies and reporting tools must mirror each other and work well together.
- To be truly effective, software-based Web-access reporting tools must produce *accurate* information.
- To be truly reliable, reporting tools must provide *metrics* that can be compared directly to quantified *standards* in the policy. ("Metrics" are quantifiable actions to be tracked, e.g., number-of-visits to objectionable Web sites.)

While attainable, the achievement of these objectives is not always easy. The paper discusses difficulties along the way and points out how to avoid them.

Note. This paper focuses chiefly on the *accuracy* of Web-access reporting systems and on the relationship of such systems to Internet usage policies. Readers who are interested in exploring the *overall* subject of policy-based reporting in more detail may wish to read other white papers on the subject. These can be found on the Web at www.wavecrest.net. Other useful information can be found on the E-Policy Institute's Web page, www.epolicyinstitute.com

Introduction/Background

General. Internet access in the work place is a double-edged sword. On one hand such access—particularly access to Web sites—can greatly increase the workforce's efficiency and productivity. After all, it facilitates useful research, provides quick answers, aids effective collaboration with colleagues and customers and enables efficient interaction with partner and supplier firms. On the other hand, Internet access can *reduce* the efficiency and productivity of that same workforce and lead to other unintended and negative consequences. It does this by tempting workers to spend time surfing on Web sites that may be loaded with interesting, entertaining, lewd or malicious content, but are not related to work at all.

> Indeed, cyber-loafing accounts for 30% to 40% of lost worker productivity, according to Framingham MA-based International Data Corporation. — *Business Week Magazine*

Enterprise managers, HR administrators and IT professionals need to configure, plan and control Web-access resources and activity.

Such casual surfing not only detracts from work force productivity, it can lead to legal liabilities. These appear primarily in the form of sexual harassment or "hostile workplace" lawsuits. Typically, such suits are filed by employees who have inadvertently or deliberately been exposed to pornographic images downloaded by other employees.

For these reasons, today's organizations need to *configure*, *plan* and *control* Web-access resources and activity in ways that optimize workforce performance while avoiding legal and personnel problems. This paper assumes that the resources are configured properly and that plans for their usage are in place. Consequently, the general focus of the paper is on *control*—more specifically on a particular *element* of control, i.e., Web-access reporting.

To explain further, Web-access control requires three things:

- Policies and standards that prescribe proper usage of Web-access resources.
- A means of monitoring Web usage—specifically a Web-access reporting tool that can gauge compliance with those policies.
- Follow-up action by management to address areas of noncompliance.

Of these three elements, the paper focuses chiefly on the second. Nonetheless, a brief discussion of policies, standards and follow-up action may provide useful background.

Policies and Standards. As indicated earlier, in most organizations Web-access resources are crucial to workforce productivity. On the other hand, they can, when misused, they can degrade productivity and lead to legal liability issues and other serious problems. For these reasons more and more businesses and other organizations are finding it essential to institute policies and standards to help govern their use. Current studies show that approximately sixty percent do so. Such organizations are finding that, to be effective, policies must:

- Be consistent with corporate culture;
- Define the types of Web-access activities that are strongly encouraged, strictly prohibited, or permissible within quantified limits;
- Contain detailed standards that are unambiguous, measurable and easily understood.
- Be well publicized and widely disseminated.
- Be balanced, reasonable and fair.
- Be designed and worded so that levels of compliance can be gauged with specific metrics and reported with high degrees of accuracy.

To be truly effective, Web-access policies, standards and followup actions must be supported by a reporting tool that produces accurate information related to the workforce's use of network resources.

Put another way, an effective policy must be closely coupled, coordinated and consistent with an associated Web-access reporting system, and it must contain standards against which Web-use activity can be accurately measured.

> To be effective, Internet access management solutions must provide quantifiable information about where users are surfing, what user or department consumes the most bandwidth, and when the peak periods tax your network.
> — *InfoWorld Magazine*

Follow-up Action by Management. Assuming that the Web-usage policy is supported with an accurate Web-access reporting system, management can use the latter's reports to determine if the workforce's activity is in compliance with the policy. If it is, further action may not be necessary. If it is *not*, management can choose from a range of options to correct the situation or effect desired improvements. These options include modification of procedures or processes and a variety of personnel actions. Among the latter are counseling, training, retraining, reassignment, reprimand, and even termination. In any case, the important point is that the information on which conclusions and follow-up action are based must be *as accurate as possible*. Inaccurate, unclear, or distorted information in reports can lead to highly counterproductive, unfair and possibly illegal actions, e.g., unwarranted reprimands or terminations, invitations to lawsuits, procedural changes that do more harm than good, etc.

The Bottom Line. To be truly effective, Web-access policies, standards and follow-up actions must be supported by a reporting tool that produces *accurate* information related to the workforce's use of network resources. We'll be exploring this subject in some detail in the pages that follow.

Web-access Reporting Software

General. This section discusses Web-access reporting software from an informational *accuracy* perspective. To preclude confusion over terminology, it begins by briefly examining the differences between *inbound* and *outbound* Web-access activity and their associated reporting tools. It then discusses the two major approaches to outbound reporting, looks at sources of the raw data from which reports are generated, and points out sources of inaccuracy that can be avoided.

Effective Web-use policies are much more concerned with visits — or human actions — than other types of hits.

Types of Web Activity and Web-access Reporting. From a very broad, high-level perspective, there are two types of Web-site activity reporting products (software). One type deals with *inbound* Web site visits, and the other deals with *outbound* visits to Web sites. Inbound activity is concerned with: "Who is looking at my Web site?" Outbound activity is concerned with "Where are my employees going on the Web?" While this paper is focused on the latter question, it may be helpful to briefly discuss inbound activity and how it differs from outbound activity.

Inbound Activity. Inbound activity consists of many users visiting one Web site. Such activity is of great interest to businesses that need or want to track incoming visits to their Web site. In this arena, the activity *of interest* is one-way (user to site), the relationship is many-to-one (many users, one site), and approaches to monitoring the visits are relatively easy to understand and implement.

Systems that *report* on inbound activity are designed to answer simple business questions, e.g., "How many people looked at *my* Web site today?" In this environment, the terms *hits* and *visits* are often used interchangeably to denote "looking at" a site. While this does no particular harm in the world of inbound reporting, interchanging these terms in the world of outbound reporting causes considerable confusion, as we shall see later.

> **Hits and Visits.** Currently there are no universally accepted definitions of the terms hit and visit. However, most IT experts define them as follows. **Hit:** A hit is any browser-related action or data display associated with Web site activity. This includes any deliberate mouse click whose purpose is to display a selected Web page in the browser. However, it also includes all the individual elements of information that appear in the browser as a result of the click, e.g., graphics, banners, ads, background audio, video images, etc. (This is the source of much confusion, as we'll see later.) **Visit**: A visit is a deliberate action (mouse click) that brings up a particular Web page or requests a particular download; a visit is *one ty*pe of hit.

Outbound Activity. Outbound activity involves large numbers of *internal* users who access and receive information from large numbers of external Web sites. This makes for a complex many-to-many relationship (many users, many sites). In addition, the activity of interest is two-way, i.e., user to site and back again. Products that report on outbound activity are more sophisticated and complex than those that deal only with inbound activity. We'll now take a deeper look at these.

> **Note**: Effective Web policies are much more concerned with visits than other types of hits. The reason is simple. Web policies deal with *human* actions and performance, and a visit is a human action. Conversely, other types of hits occur automatically and don't reflect human behavior.

Outbound Web-access Reporting. One way or another, outbound Web-access reporting software measures, characterizes and depicts a workforce's Web-site visitation activity. Ideally, the software answers such questions as these: "Which users visited which sites, when did they do so, how often did they do so, what type of content were they seeking, how much bandwidth was consumed in the process, and were the visits in compliance with our usage policy?" If the software does this well, it provides reports that *accurately* represent or model Web-visitation activity on the part of a single user, a group of users, or an entire enterprise.

Used chiefly to gauge compliance with policies and standards, outbound Web-access reports *primarily* but not exclusively address two issues:

- what types of sites are individual users or groups of users visiting?
- how much activity are the users engaged in?

Let's see how this is done.

> **Note**: As stressed throughout this paper, the chief purpose of outbound Web-access reporting tools is to help gauge compliance with Internet usage policies. However, it's worth noting that such tools can also be used for other

purposes. For example, they can help a) assess work force productivity, b) evaluate bandwidth usage, and c) develop prorated departmental "charge-backs" for network resource usage. Because this paper focuses primarily on Web *policy* issues, particularly the accuracy of enforcement information, these other three subjects are not explored in detail herein. This does not imply that they are somehow unimportant.

URLs and time stamps are crucial to accuracy in Web-access reporting, but they both have certain limitations that need to be understood.

One way or another, most outbound Web-access reporting products use pre-existing log files as their source of raw data. Log files are essentially tables of highly detailed electronic records (time-tagged logs that provide a running history of outbound Web-site visitation activity). Kept in proxy servers, firewalls or caching appliances, they list *all* hits associated with *all* outbound activity. Log files time-stamp each hit, identify visitors, identify URLs (Uniform Resource Locators), count bytes, etc. The reporting products then produce reports by analyzing and processing the following data points:

- **User ID** (login name or anonymous IP address)
- **URL** (Web site name and type, e.g., HTML, GIF, audio, etc.)
- **Time-Stamp** (the date/time associated with every hit).

The first of these, user ID, is easy enough to understand and will not be addressed in any detail here. The latter two, however, are crucial to understanding the accuracy issues related to Web-access reporting products. URLs and time-stamps are highly useful, but they both have certain limitations that need to be understood. Let's see what they can and cannot do to help provide accurate reports reflecting the types of sites visited and the levels of Web-access activity.

Using URLs to Determine the *Type* of Activity. By itself, a URL (Uniform Resource Locator) has very little if any meaning to non-technical personnel. For a URL to be meaningful, i.e., to help determine *types* of sites visited, the software must correctly *categorize* it (sports, pornography, finance, etc.). The software must also classify the URL as either a visit or other type of hit. If the URL reflects an actual visit, the software should also be able to label it as *acceptable* or *unacceptable* in accordance with the organization's Web sites rating policy. Some products do these things effectively; others do not.

Using Time-Stamps to Determine the *Magnitude* of Activity. Managers and administrators need an accurate picture of *each* user's level of activity. They need this information to fully evaluate compliance with acceptable use policies, identify serious problem areas, and verify that the most productive sites are being fully exploited. Compliance with policy is particularly important. For example, in a given category of sites (e.g., sports), a modest level of activity may be perfectly acceptable while an excessive level may be unacceptable; the reporting system needs to make this distinction very precisely. (This is analogous to a telephone usage policy that allows a *reasonable* number of personal calls.)

Some products claim to measure and depict the exact amount of time the user spends on the various types of Web sites visited. This is an appealing but potentially very misleading notion.

To gauge the magnitude of outbound Web-access activity, competing products generally use time-stamps in one of two different ways. Let's look at them.

1. On-Site Time. Some products claim to measure and depict the exact amount of time the user spends on the various types of Web sites visited. This is an appealing but potentially very misleading notion. It's appealing because managers and administrators are accustomed to dealing with *time*. They frequently ask questions that have a *duration* dimension. For example, how *much* time is required to do a certain job, and how *much* time do specific employees spend on specific tasks?

It would of course be very useful if on-site time *could* be accurately measured. However, that's not possible. The time-stamp only indicates the *instant* at which a Web page was accessed, or the moment when an item of content was displayed. It says nothing about the duration of time the user spent actually viewing the site. After all, the user could have accessed a Web page and then received a phone call, gone to lunch, attended a short-notice meeting, been interrupted by a colleague, or gone to the restroom. Because unpredictable human actions are involved, the reporting software simply cannot determine with any certainty what occurred between log entries or time stamps. When the user finally clicks on another URL, the logfile takes note of this and, in a hypothetical sense, assumes that the user has

been actively viewing the first site the entire time. If, as some systems do, the reporting system generates a report on the basis of this assumption, the manager or administrator who reads the report can easily draw erroneous conclusions.

Compounding the problem further, products that claim to provide accurate measurements of on-site time typically label all Web-site activity as hits. Along with the user's deliberate clicks (visits), they include the various images and sounds that appear automatically as a result of the click, e.g., banners and ads. Such mixing of visits with other types of hits seriously distorts the true nature of the user's activity, particularly the *level* of activity. That is, the act of reporting all activity as hits gives the false impression that the user has been much more active than is actually the case.

The following comparison illustrates this point. It shows two different users visiting two different Web pages, once each. (The users' names are fictional, but the URLs are real.)

1. John Doe visits http://www.whitehouse.gov/text/index.html . This is a text-only page, i.e., no images, banners or ads. The logfile registers only one hit. In this case, one hit equals one visit.

2. Mary Smith visits http://www.cnn.com. This is a complex page with 22 images, banners and ads. The logfile registers 23 hits, i.e., one click and 22 data items. In this case, 23 hits equals one visit, but it's not obvious.

Approximately 75% of all hits are not true visits. The most effective products count and report the number of times the user clicked on specific Web sites and Web pages.

Let's assume that John and Mary are being monitored by a reporting system that fails to distinguish visits from other types of hits. Mary would appear to be much more active than John, even though the amount of user-initiated activity is the same in both cases. If the two sites happened to be "unacceptable" Mary could be considered much more abusive than John and possibly be subjected to harsh but unwarranted disciplinary action.

With respect to this issue, it is particularly important to note that approximately 75 percent of all hits are not true visits.

> **Note**: The magnitude-measurement problems discussed above are usually associated with products that use software concepts and techniques originally developed for use in tools that report on *inbound* activity. When dealing with inbound activity, these concepts and techniques may be perfectly satisfactory, but they are far too simplistic to handle the more sophisticated requirements of outbound reporting.

2. Visit-Counts. The most effective products *count* and report the number of times the user clicked on (i.e., visited) specific Web sites and Web pages. Such products clearly distinguish true visits from other types of hits. This prevents the distortions and inaccuracies seen in reports that make no such distinction. Visit-counts constitute a clear metric for measuring compliance with policy standards that prescribe acceptable levels of activity.

To supplement the visit-counts, such products also calculate the minimum time it takes for an average Web page to load in the browser. Although not perfect, this is the most conservative and accurate approach to estimating the time and bandwidth used.

It is critically important for managers and administrators to ensure the accuracy, reliability and relevance of the informaiton they are using to implement and administer Web-use policies.

Compared to the "on-site time" approach, a visit-count can provide a much more accurate and reliable representation of a user's *intent* while online and the level of his or her activity. Simply put, visit-counts reflect the user's actual actions, which is the subject of the policy in the first place. And they do this in a way that doesn't obscure the most valuable information—number-of-visits—with clutter that's traceable to automatic, machine-originated events such as the appearance of unsolicited banners and ads.

Summing up, it is critically important for managers and administrators to ensure the accuracy, reliability and relevance of the information that they use to implement and administer Web-use policies. This can be achieved with the help of basic understanding of differences between: a) inbound and outbound activity, b) hits and visits, and c) the two approaches to gauging the magnitude of Web-access activity. Hopefully, this paper contributed to that understanding.

Conclusions and Summary

Web-use policies employing clear standards can help organizations ensure productivity, control costs, and avoid legal liabilities. And if designed and administered properly, they can do this without hurting morale. However, to obtain these benefits, management needs to couple the policies with well-designed reporting tools. The policy and the reporting system should use similar terminology, and the standards in the policy should refer to the same parameters and units of measure as the metrics in the reporting system. When this is done, the result will be reliable metrics and accurate reports. These in turn will provide accurate indications of compliance and help ensure sound decisions and productive actions.

Remember: reporting systems will produce inaccurate results if they employ software techniques and concepts developed for inbound systems. Effective reporting systems are specifically designed for outbound reporting.

At a more detailed level, it's important to note that metrics used for policy enforcement must actually be *obtainable*. That is, the raw data from which they are derived must contain the basic information to support the metrics in the first place. To illustrate this point, certain traditional metrics such as time-measurement don't work in the Internet world. The raw data simply isn't available in the log files. On the other hand, counting mouse clicks is a reliable, achievable and accurate method of gauging the *level* of Web-access activity.

Different Web-access reporting tools provide varying levels of accuracy. Some are quite accurate, and others are quite *inaccurate* (although obviously they don't reveal this in their advertising). To select and use Web-access reporting tools that best meet the organization's needs, business managers and HR professionals should be familiar with certain confusion factors, e.g., the:

- failure of some products to differentiate visits from other types of hits
- limitations of log files
- difference between inbound and outbound reporting.

They should also be aware that outbound reporting systems will produce inaccurate results if they employ software techniques and concepts developed for inbound reporting systems, e.g., treating all Web-access activity (hits) alike.

In summary, effective outbound reporting systems:

- Are designed specifically for outbound reporting.
- *Clearly* distinguish visits from other types of hits.
- Use "visit-counts" as the primary metric to determine levels of activity.
- Provide estimates of download time as a secondary indicator of level of activity.
- Don't claim to measure and display time-on-site with any degree of accuracy.
- Categorize *all* types of Web-access activity, not just unacceptable visits.
- Provide metrics, i.e. measurements and information, to which policy standards can be directly compared, e.g., number of visits allowed in different categories.
- Can be used as the basis for *development* as well as implementation of Web usage policies.