Is Internet Advertising ready For Prime Time?

Xavier Drèze,

and

Fred Zufryden*

April 5, 1998

ABSTRACT

Advertising on the World Wide Web is growing at a fast pace. However, it is difficult to compare advertising effectiveness on the Internet relative to standard media, such as broadcast and print, because current measures of advertising effectiveness on the Web are not standardized and incorporate significant measurement errors. In this study, we investigate issues relating to the accurate measurement of advertising GRPs, Reach and Frequency on the Internet. Moreover, we suggest critical measurement issues that need to be resolved before Internet advertising can be considered as an integral part of a company's media mix.

*Xavier Drèze is Assistant Professor of Marketing and Fred S. Zufryden is the Ernest W. Hahn Professor of Marketing, both at the Marshall School of Business of the University of Southern California. The authors would like to acknowledge and thank Paul Grand, Chairman of Netcount, L.L.C., for providing the data used in this study.

INTRODUCTION

Online advertising revenues reached $906.5 million in 1997 (Internet Advertising Bureau, 1998). Although advertising expenditures on the Internet remain relatively small in contrast to standard media, this figure is approaching that of outdoor advertising and represents nearly a three-fold increase over the last two years. In 1995, only $312 millions were reported on Web advertising in contrast to 38.1 billions for the television medium (Jupiter Communications, 1996).

The spectacular growth trend in online advertising revenues, that has been experienced so far, is expected to continue. Thus, it has been projected that Internet advertising revenue will reach $ 1.9 billion in 1998, grow to $ 3 billion in 1999, and reach a level of $ 4.3 billion in year 2000 (Jupiter Communications, 1998). Indeed, the latter growth statistics suggest that the Internet is quickly becoming a very significant advertising medium. Moreover, a substantial increase in revenues, from $1.5 billion in 1996 to a level of $3.7 billion in 1998, has been noted for expertise, access, software, content and commerce on the Internet (Advertising Age, 1998). Thus, the concurrent growth of electronic commerce is expected to further fuel the growth of advertising activities and revenues on the Internet in the future.

Along with the growth of electronic commerce and advertising on the Web, the number of Web users has also been growing rapidly (Electronic Advertising and Marketplace Report, 1997). The number of Web users has been estimated at 36 million in 1996 and is expected to grow to 170 million users by the year 2000 (I/PRO, CyberAtlas, 1996). All of these factors suggest that Web-based advertising will ultimately become a very important component of a company’s media mix. Already, numerous companies are committing large budgets to advertising on the Web. For instance, “Amazon.com,” an online book retailer, has recently signed a contract for $19 millions with America Online to rent space on the latter’s home page to feature its banner ads for the next three years (see Online Money News, 1997).

Internet advertising has now reached the point at which many companies are considering it as being a viable alternative to traditional media. However, to render the Internet fully viable as an advertising medium, companies need standardized measures (such as Reach and Frequency) that will allow them to compare advertising effectiveness on the Internet to that of other media. Unfortunately, this is where a number of major measurement shortcomings currently exist on the Web. This is because Web servers typically provide statistics that can tell advertisers how many pages were requested, how much time was spent on each Web page, and what types of computers made the page requests. However, it is generally very difficult to translate the page requests into specific audience viewing behavior estimates. In particular, due to the present problems associated with identifying unique visitors to a site, it is difficult to accurately measure the impressions, Reach and Frequency of banner ad exposures for a target audience. Thus, the fundamental questions of “How many people visit a Web site?” and “What types of people visit a Web site?” are generally unanswered by current Web-based measures. Hence, it is difficult to compare advertising effectiveness on the Internet relative to that in media such as broadcast and print. Although standardized measurements exist for the latter media, standards have not yet been formalized for assessing advertising effectiveness on the Web. Until these standards emerge for the Internet, we argue that Internet advertising will not be truly “ready for prime time” in the sense that its effectiveness will not be comparable to the advertising effectiveness of other media. Hence, advertisers will not be able to justify shifting significant advertising budgets away from traditional media and allocate them to online advertising. This means that it is important to provide accurate measures of *impressions* as well as *Reach* and *Frequency* of banner ad exposures relative to target audiences on the Web.

The purpose of the study is to determine the nature and magnitude of the errors that exist in current Web-based advertising effectiveness measures. We evaluate the accuracy of current methods in measuring Reach, Frequency and Gross Rating Points (GRP) for banner ads on the Web. We also provide guidelines that will allow for a better interpretation of results from current measures and enable more meaningful comparisons to be made between advertising effectiveness on the Web relative to standard media.

MEASURING ADVERTISING EFFECTIVENESS ON THE WEB

Measurement Problems

At the present time, despite ongoing efforts towards this end, there does not appear to be any widely accepted measurement standards for advertising on the Web. Third party companies, such as Netcount and I/PRO have proposed Web-specific measures such as click-through rates (i.e., clicks on banner ads on a publisher's Web page that result in the transfer of a surfer to the advertiser's Web site) and ad transfers (pages containing an advertiser's advertising message downloaded to a surfer's PC), developed from server log file records, to assess the effectiveness of banner ads within the Web-based multimedia environment. Interestingly, recent empirical evidence has suggested that the use of click-through rates is likely to undervalue the Web as an advertising medium (Briggs and

Hollis, 1997). In contrast, to the aforementioned third-party census-based measurement procedures, companies such as MediaMetrix or Milward Brown Interactive have developed market measurement methods based on a *sample* of home-based PCs in the United States (see Coffey and Stipp, 1997). Despite the advantages of the latter data source for evaluating individual visitor behavior on the Web, a potential limitation of online panel data is the sampling bias due to the omission of work- and school-based PCs from their sample. Another problem is that of representativeness. The number of panelists required to accurately reflect the behavior of Internet surfers is far larger than the number of households required to track fifty TV channels or one hundred cereal brands. Some researcher (e.g., Drèze, Kalyanam, and Briggs, 1998) are working on methods that improve the representativeness of these online panels. However, these methods are still at the prototype stage.

Some Web-based companies have taken steps to provide Reach and Frequency measures on the Web in an effort to provide comparability with standard media (I/PRO-DoubleClick, 1997). However, the accuracy of these measures is limited by the current measurement problems that exist on the Web. In particular, there are two essential measurement problems that may bias such measures in connection with the measurement of the effectiveness of banner ads on the Web:

1. *The problem of identifying unique site visitors on the Web* – In traditional media, advertising effectiveness measurements are usually collected through some kind of survey or panel. In these surveys or panels, potential viewers or listeners are uniquely identified (by a phone number or household address) and can, if needed, be tracked over time. Using this procedure, each record made can be linked to its originator

without ambiguity. On the Internet, measurement companies have done away with costly surveys and can track users' actions from the log files of the sites they visit. Measurements of visitor traffic and flow patterns to, from, and within a given site are generally established on the basis of the visitors' IP (Internet Protocol) addresses. Unfortunately, these addresses may not be uniquely assigned to visitors by their Internet Service Provider. For example, several Internet users can be assigned the same IP address in multi-user systems such as America Online. In addition, from one session to another, Internet users who use Internet Service Providers that use dynamic IP allocation may have different IP addresses assigned to them. To complicate matters further, users can be assigned multiple IP addresses even within a single Internet session by Internet Service Providers that uses multiple "proxy servers[1]." All of these problems make it difficult to accurately link the actions recorded on a web site's log file to *unique* visitors of that site. Consequently, these problems may seriously affect the accurate measurement of advertising effectiveness measures such as ad Reach on the Internet.

The following provides four alternative scenarios that illustrate the potential problems encountered when IP addresses are used to measure visitor behavior:

a) Visitor 1 accesses the Internet from two different PCs:

IP1 $\rightarrow$ Visit 1

Individual Visitor 1

IP2 $\rightarrow$ Visit 2

---

[1] A proxy server is a centralized computer that consolidates requests from multiple client computers. E.g., if 20 different AOL users request the same MSNBC page, the AOL proxy server will make only one request to MSNBC and forward the answer to the 20 originating computers.

b) Visitor 2 makes one visit but uses an ISP with proxy servers. His requests are handled by two different proxy servers (with two different IP addresses) during the same visit:

IP3 $\rightarrow$ Visit 1

Individual Visitor 2

IP4 $\rightarrow$ Visit 1

c) Visitor 3 makes two visits and is assigned a different IP by his ISP at each visit:

IP5 $\rightarrow$ Visit 1

Individual Visitor 3

IP6 $\rightarrow$ Visit 2

d) Visitor 4 makes two visits and uses the same ISP as user 3. For his second visit, he is assigned the same IP address as used by visitor 3 on his first visit:

IP7 $\rightarrow$ Visit 1

Individual Visitor 4

IP5 $\rightarrow$ Visit 2

Looking at the log file of the site visited by our 4 individuals, we would erroneously conclude that 7 different individuals visited the site. We would also believe that visit 1 of visitor 3 and visit 2 of visitor 4 were made by the same individual.

2. *The problem of caching* – An important determinant of the measurement of banner ad effectiveness is the number of pages requested by a surfer on the Internet. A Web page is a document that may contain text as well as graphics, sound and video files. A Web page thus provides a basic building block for a Web site, which will typically

be comprised of one or more inter-linked Web pages. Consider an individual surfer (Jane) who requests a page (A) from a Web Site. This page contains a series of links to other pages as well as a banner ad for a car manufacturer (Ad 1). After reading the page, Jane follows one of the links to another page (B). Upon reading this second page, she clicks on the "Back Button" to go back to the first page (A). This results in a second exposure to the ad. Unfortunately, in an attempt to speedup the information flow, Jane's browser will not make a second request for page one (and the car ad), but will rather retrieve the information (both page A, and Ad 1) from a local cache (typically on the surfer's PC hard drive) [2] where it stores pages after receiving then. This means that the web site's server will not record Jane's second or any of her subsequent exposures to the banner ad. Consequently, caching seriously biases ad effectiveness measures, such as impressions and exposure Frequency, for banner ads. In fact, our study revealed that an average of 30% of page requests are not recorded by servers, either because the Web pages were cached on the surfer's PC or by "proxy servers" somewhere between the server and the surfer. Although the specific problem of caching is not encountered in traditional media, similar problems have plagued the measurement in standard media as well. For example, the time-shifted viewing of television programs by means of video recorders also lead to understatements of viewing behavior for the TV medium.

A third problem affecting the reliability of reported measures lies in the fact that there is a difference between requesting a page and actually reading it, or even receiving it. A

[2] Caches are temporary storage areas in which a computer system stores information that may be too

lot of things can happen between a user request and the actual processing of the information contained in that page by the user. He/she can decide that the request takes too long to complete and cancel it. He/she can be attracted by a message displayed at the top of the requested page and request another page without reading all the information contained on the first one. The page might even be too long to fit on the screen, and the user does not that the time to scroll it to view its entirety. This type of problem plagues any medium. Newspaper subscribers do not necessary read a newspaper in its entirety. When watching TV, a viewer could be distracted by the telephone or go to the kitchen and miss a particular commercial advertising. Obviously this problem will tend to bias reported measures upward. Although we recognize this problem, we do not specifically consider it in our study in view of the associated measurement difficulties.

Defining and Measuring Ad Effectiveness on the Web

Consider the measurement of advertising effectiveness for a banner ad on the Web. The currently favored Web measures are Page Views and Click-through rates. However, to insure comparability with standard media, the following standard measures need to be defined and measured for banner ads on the Web:

1. *Reach* – The net unduplicated number, or percentage, of a target audience that have had opportunity to see a banner ad one or more times.
2. *Frequency* – The average number of times an individual has had the opportunity to see a banner ad (conditional on the fact that he/she has been exposed to the banner ad at least once).

---

lengthy to retrieve from their original location and that it may need in the near future.

3. *Gross Rating Points* – This measure, also called impressions, is the sum of all potential exposures to a banner ad without accounting for audience duplication. GRP may also be defined as the product of Reach and Frequency (i.e., GRP = Reach * Frequency).

As noted earlier, the development of advertising effectiveness measure, such as Reach, require that the exposure patterns of individuals be known. This means that we must be able to identify and differentiate *unique* individual Web site visitors. Currently IP addresses of visitors are the most common means of identifying unique Web site visitors. Consequently, the number of distinct addresses that access a given page are used to compute the Reach of a banner ad on the Web. Unfortunately, because of the non-uniqueness of IP address assignment to Web users that has been described above, the resulting Reach figures are severely biased.

Likewise, the development of effectiveness measures, such as Frequency, requires that the complete exposure patterns to banner ads of unique individuals be known. This measurement is also complicated by the problems of identifying unique visitors and those of caching that have been described above.

Study Methodology

Our study focuses on the evaluation of the nature and magnitude of errors that now exist in measuring Reach, Frequency and GRP for banner ads on the Web with current measurement methods. We proceeded according to the following four study steps: (1) We first examine five web sites that track their visitors by the means of unique visitor IDs to estimate the bias created by the use of IP addresses as a means of visitor identification; (2) In the next study, we explore the implications of study 1 for measuring advertising effectiveness measures, including GRP, Reach and Frequency; (3) next, we evaluate the effectiveness of three methodologies that attempt to recover cached pages; (4) last, we set up a test site to evaluate the impact, on surfers, of using procedures to uniquely identify them and to prevent caching.

Our approach (in study 2) involves the computation of Reach, Frequency and GRP by means of current measurement procedures. That is, the Internet users' IP addresses are used to identify unique visitors to a Web site. At this stage, the recording of page requests from Web site servers are used to measure potential exposures to banner ads without accounting for potential repeat exposures to pages that may have been cached on individual surfers' PCs.

In order to examine the magnitude of the errors inherent in current measurement methods, we examined various web sites that use unique visitor IDs. The visitor IDs were generated using a combination of cookies and user passwords[3]. These visitor IDs allow us to compute actual measures of Reach, Frequency and GRP that accurately accounted for unique visitors.

We gathered data from five Web sites. These Web sites had widely different target audiences (education, news, entertainment, and information database) as well as sizes (1,000 to 90,000 visits per site). Our methodology involved a comparison of statistics generated using only IP address information (as would generally be the case for current measurement methods), and an IP address analysis supplemented by information

[3] One must note that this tracking methodology is likely to impact the behavior of visitors (see study 4 for an analysis of this impact). However, the purpose of this study is not to describe the behavior of the Internet population, but rather to evaluate how accurately one can measure this behavior. Hence we do not see this as a significant problem.

from visitor IDs (to obtain contrasting estimates of the actual levels of the advertising effectiveness measures).

The third stage of our research evaluates the effectiveness of three cache recovery strategies. Three techniques are currently used to measure the requests of cached pages. (1) Do nothing, (2) Full Path Recovery, and (3) Partial Path Recovery. We test the ability of these techniques to produce accurate statistics of ad exposure Frequency (caching does not affect Reach measures).

The purpose of our last study (study 4) was to evaluate the impact of implementing procedures for better monitoring Web surfers. Specifically, does the use of Cookies[4] as visitor tracking mechanism or the implementation of a cache-defeating scheme alter surfers' behavior? To study this question, we modified an existing site to implement a 2x2 experimental design (see Figure 3). In this setup, half of the site visitors were issued Cookies and the other half were not. The second (and orthogonal) treatment was to issue an immediate expiration date on the pages served to half the visitors. This would force the browsers of these visitors to reload the page from the server each time they want to access it, instead of retrieving them form their cache). Because of implementation issues, we were not able to test whether the use of passwords significantly impact the behavior of online surfers.

STUDY RESULTS

Study 1 - Current Biases in Estimating Number of Visitors and Visits.

[4] Cookies are small strings of text that are sent by a web site to a visiting surfer's PC, stored on that PC, and retrieved by the web site on subsequent page request (i.e., these requests are tagged with the cookie).

We first investigate the biases inherent in current measurement methods, which rely on IP address identification, in connection to the estimation of number of visitors and visits. Using Visitor ID information, we find an average ratio of 1.16 IP addresses per visitor. This suggests that there is a significant problem in using IP addresses as a measure of number of unique visitors as different IP addresses were assigned to the same visitor in 16% of the address assignment instances. We also find an average ratio of 2.1 visitors per IP address. This suggested that a given address is assigned on the average to approximately 2 visitors. Moreover the non-uniqueness of IP addresses is further supported by the fact that the average number of visits with multiple IP addresses (within a visit) was 3.5%. These represented cases where multiple proxy servers were used to satisfy a given visitor's requests within a given Internet session.

To assess the importance of using the proper procedures to identify users, we first compute basic descriptive statistics for each of the five sites. First, by using visitor IDs as the means of differentiating users. Second, by using IP address for that purpose. The summary results are shown in Table 1 (complete results are shown in Appendix A). As one can see, using IP addresses as a mean of identification will lead to an underestimation of the number of visitors (39% on average) as well as number of the true number of visits (35% on average). Naturally, underestimating the number of visitors and visits leads to an overestimation of the number of pages seen by each visitor (64%) as well as the time spent during each visit (79%). Finally, the reallocation of IP addresses to multiple users leads to an overestimation of the number of repeat visits of about 9%.

TABLE 1: Measurement Error

Measurement	**Average Error**

# of Visitors	-39%[5]
# of Visits	-35%
# Pages accessed / Visit	+64%
Time Spent / Visit	+79%
# Repeat Visits	+9%

Study 2 - Implications for Measurement of Advertising Effectiveness

Next, we investigate the accuracy of IP based methods when measuring Reach, Frequency, and GRP. We look at banner ads appearing in two of the five Web sites that we studied (the other sites did not display advertising). Specifically, we studied three ads from site 1 and two ads from site 2. For the purpose of this test, we ignore the impact of caching on GRP measures. The issue of caching is studied in detail in the next section. At this stage, we only consider biases induced by visitor identification. The results of this investigation are summarized in Table 2 (detailed results appear in Appendix B).

TABLE 2: Errors in Measurement of Advertising Effectiveness
Based on IP Address Identification

		% Error Range	
	Average % Error	**Low**	**High**
Reach	+25%[6]	+12%	+42%
Frequency	-1.1%	-13%	+9%
GRP	+22.7%	+22%	+23%

As noted in Table 2, the measurement of advertising Reach is subject to the greatest error magnitude, with an average error of +25% and an error ranging from +12% to as high as +42%. Frequency was least subject to error by current measurement methods with an average of –1.1%. Here, the error ranged from a low of –13% to a high

[5] -39% means that IP based measures underestimate the true number of visitors by 39%.

of + 9%. Finally, GRPs were shown to result in an error of + 22.7% with a variation over a narrow range of +22% to +23%. There is less variance in GRP error than in the two other measures because the error on Reach and Frequency are negatively correlated. Indeed, the numerator of the Reach formula and the denominator of the frequency formula are the same (number of people having seen the ad). Since GRP is equal to the product of Reach and Frequency, the numerator and denominator cancel each other. In fact, the only source of error in the GRP formula comes from the estimation of the total number of visitors (assuming that caching is not an issue). This means that, within a site, the error in GRP is *constant*. It is equal to the inverse of the error in unique visitor count. For instance, using IP addresses to measure GRP, we overestimate Site 1's measures by 23.09%. This is due to the error in visitor identification of –18.76% (1 / 0.8124 = 1.2309). Similarly, the errors for Site 2 are 22.14% and –18.12% for GRP and visitor identification (1 / 0.8188 = 1.2214). In short, if one can estimate the error rate in using IP addresses to identify users for a given site, one can correct all GRP measures for that site.

The Frequency distribution of exposures often provides a useful media analysis tool for studying the impact of a given media schedule on a target market. Several different model approaches have been proposed in the literature to measure the Frequency distribution of exposures (e.g., see Metheringham, 1964; Headen et al, 1977; Zufryden, 1987; Pedrick and Zufryden, 1991). The results of Table 2 suggest that the Frequency of exposures distributions, that are developed from current measurement methods, may be subject to significant errors.

To illustrate the specific nature of the potential errors in exposure distributions that

---

[6] i.e., IP based measurement overestimate true Reach by 25%.

may be produced by the current methods, we used the commonly applied Beta Binomial Distribution (BBD), first proposed by Metheringham (1964), as a model of the distribution of banner ad exposures over a target population. Here, this model is used to describe $P(i|N;m,n)$, the probability that a Web surfer is subject to i banner ad exposures given a maximum of N exposures, and may be stated mathematically as:

$$P(i|N;m,n) = \frac{N!}{(N-i)!i!}\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\frac{\Gamma(m+i)\Gamma(n+N-i)}{\Gamma(m+n+N)} \qquad (1)$$

for $i=0, 1, 2, \ldots, N$ banner ad exposures, with $m, n > 0$ model parameters.

We considered the exposure patterns to a banner ad on a given Web page based on measurements of Reach = 0.75 and GRPs = 1.75 from current methods by using only IP address identification. Then, we obtained estimates of actual Reach and Frequency for the banner ad by applying the average % error corrections developed in Table 2. Thus, the actual Reach and GRP figures were estimated as 0.56 and 1.35 respectively. We then fitted BBD distributions to the corresponding Reach and GRP figures by a method of means and zeroes (e.g., see Pedrick and Zufryden 1991). Figure 1 illustrates the distributions based on actual and estimated Reach and GRP values that were so developed for a case situation where the maximum number of banner ad exposures in the target market is N=16. Here, we note that the exposure distribution estimated from current measurement methods resulted in frequencies that overstated the actual frequencies over intermediate values of exposure levels and then slightly understated actual Frequency values for higher levels of exposures. Thus, Figure 1 suggests that errors in Reach and GRP may create significant skews and consequent biases in the resulting patterns of distributions of exposures.

==================

Insert Figure 1 about here

==================

Study 3 – Cache Recovery Algorithms

We now turn to the implications of caching. As we have mentioned earlier in this paper, the purpose of caching is to increase the apparent speed of data transfer between a user's computer and the site he/she is visiting. This is done by saving some information on the user's PC and accessing that information more speedily instead of the web site when it is needed. As a result, web site logs will record only one exposure per visit from any visitor (assuming that the visitor's caching algorithm works perfectly). This is a problem when measuring GRP since repeat exposure to an ad during a single visit will not be recorded. Only repeat exposures across multiple visits or sites are.

To measure the impact of caching on GRP measures, we modified an existing web site so that it would send a special instruction, along with each page, that would disable caching on the user's computer. This yielded a log file with a complete record of all page accessed. We then simulated the impact of caching by keeping only the first exposure of each of the pages during each visit (i.e., Perfect Cache). Lastly, we applied three different cache recovery strategies to this file and compared various statistics obtained after cache recovery with the statistics obtained on the raw (no cache) file. The three strategies, described below, are: (1) No Path Recovery, (2) Full Path Recovery, and (3) Partial Path Recovery.

==================

Insert Figure 2 about here

==================

To illustrate these strategies, consider a Web site, which consists of a tree like structure linking its pages, defined as nodes on the network (A to G), as shown in Figure 2. Consider now the situation of a surfer who visits the site in the following way: He/she starts at A, then goes to B, D, and G. At this point our user presses the “Back Button” on the browser to go back to D, then goes to H. Now, he/she backs up through D and B, visits E, backs down to B and goes to F. At this point, the user decides that he/she has gone deep enough and uses the browser’s Go Menu to jump back directly to the starting point A, and then finishes the visit in C. The total visits yielded 13 page exposures. Nevertheless, only 8 page requests would have been made.

Possible algorithms for inferring page requests and thus recovering the sequence of cached pages include the following:

a) *Partial Path Recovery* - When the server notes a user jumping from page H to page E, it can infer that the last jump was made from B to E, and thus a dummy hit to B is added. This is a safe correction to make since along with a page request, a browser is sending information about the page from which this request originated.

b) *Full Path Recovery* - A bolder approach is to infer the shortest path from page H to E. In this case: H-D-B-E. And then account for the two unseen pages by adding dummy hits D and B.

c) *No Path Recovery* - Make no attempt to recover cached pages.

Before looking at GRP measures, we can look at a more basic statistics of site visits: the number of page accessed per visit. Table 3 summarizes the results that were obtained in measuring the average number of pages per visit for no cache, perfect cache (i.e., No Path Recovery), full path recovery and last page recovery algorithms.

TABLE 3: Implications of Caching Alternatives on the Measurement of Average Number of Pages Accessed Per Visit*

	No Cache	**Perfect Cache (No Recovery)**	**Full Path Recovery**	**Last Page Recovery**
Average No. of Pages per visit	2.63	1.63	1.87	1.64
Accuracy	100%	62%	71%	62%

* Based on a sample of 627 visits.

As may be noted in Table 3, we found that 38% (100% - 62%) of the pages in our test case were cached. Moreover, the full path recovery algorithm recovered 24% (9% / 38%) of the cached pages while the last page recovery algorithm recovered less than 1% of the cached pages.

We can now turn to the impact of cache recovery on GRP measures. We selected two ads that were running on the server during the test and computed GRP measures for the No Cache situation as well as for each of the cache recovery strategies (See Table 4). As might be expected from the results of Table 3, the last page recovery algorithm did not yield much improvement over the No Recovery option. For the first of the two ads, the full path recovery algorithm recovered 40% of the missing pages (the amount of error decrease by 40% from 50% to 30%). For the second one, it did not recover any.

TABLE 4: Path Recovery Effectiveness

		No Cache	No Recovery % Error	Partial Path Recovery % Error	Full Path Recovery % Error

A	Reach	0.20	0.20	0%	0.20	0%	0.20	0%
D	Frequency	2.75	1.38	50%	1.39	49%	1.92	30%
1	GRP	0.54	0.27	50%	0.27	49%	0.38	30%
A	Reach	0.19	0.19	0%	0.19	0%	0.19	0%
D	Frequency	1.41	1.05	26%	1.05	26%	1.05	26%
2	GRP	0.27	0.20	26%	0.20	26%	0.20	26%

Despite the significant improvements obtained for the first Ad with the full path recovery algorithm, we do not favor the use of this approach. This is because there are problems with its use. For example, if users do not take the shortest path between one page to the next, then the path recovered will be incorrect and thus this will lead to missing Web pages (i.e., only Partial Page Recovery). In addition, if a surfer uses his/her browser's "Go" menu to jump back to the last page, then the path recovered will contain pages that were not seen a second time and extra pages will be recorded (i.e., over correction). By looking at each of the 627 visits analyzed in Table 3, we can compute the number of times that the Full Path Recovery method (a) does not recover any cached pages, (b) recovers only a fraction of the cached pages, (c) recovers the exact number of cached pages, or (d) over-corrects the number of cached pages. Table 5 shows the Frequency of occurrences for these possibilities.

TABLE 5: Full Path Recovery Statistics:

	Occurrences (%)
No Page Recovered	58%
Partial Recovery of Pages	23%
Perfect Recovery	0%
Over-Correction	19%

In particular, Table 5 suggests potential problems with full path recovery. Although Full Path Recovery improves the numbers on average, we note that it often overcompensates (19% of the time), and consequently leads to estimates that are actually less reliable than the raw numbers. As shown in Table 6, even though the overall imprecision (Mean Percent Error) of the measures is greater when using Full Path Recovery than when using raw data, the reliability of these measures (Standard Deviation) suffers greatly.

TABLE 6: Reliability of Measures:

	Mean Percent Error	**Standard Deviation**
No Recovery	19%	0.20
Full Path Recovery	6%	0.33

Study 4 - Implications of Visitor Monitoring

The previous studies showed that the overall accuracy of some specific current methods leaves a lot to be desired. However, they also show that with appropriate monitoring instruments in place one can remedy this lack of precision. In our case, the monitoring that needs to take place relates to two areas: first, an accurate tracking of visitors needs to be performed, and second, one needs to be able to account for cached pages.

The monitoring of visitors can be performed in a variety of ways (e.g., passwords, digital certificates, cookies, etc.). Passwords are widely used by commercial sites that have a membership element. Companies such as E*Trade or The Wall Street Journal require their users to be registered and to use a password to access their sites.

Digital certificates provide third-party certification of a user's or Web Server's identity. They are provided by an entity that both the Web Site and the users trust (such as VeriSign). The digital certificates are unique to the individuals to whom they are issued, and are encrypted to prevent tampering. They are used by such companies as K-Mart or the Virtual Vineyard.

Cookies are much simpler devices than either digital certificates or passwords. As noted before, cookies tag requests from individual users (or more accurately, the requests from the computer used by the surfer). With this simple mechanism, web site operators can tag individual users with unique cookies and then track their progress through the web site or their subsequent site visits. Thus, ookies provide a simple and effective way to track users. However, they do not provide any information about who the users are as do digital certificates or the registration processes associated with requesting a password would. In addition, cookies may have other measurement limitations. For example, cookies cannot distinguish multiple users of one computer (except for multi-user operating systems such as Unix). Furthermore, Web users who may be sensitive to privacy issues could delete the resident file on their hard drive which stores the cookies, and thus erase their previous viewing history. Indeed, although benign in nature, Cookies have received a lot of negative press with respect to the issue of invasion of privacy and have been the subject of controversy (e.g., Komando, 1997).

Monitoring cache access is more straightforward than tracking individuals. However, it also interferes more with their actions. The simplest way to deal with caching is to disable the caching of pages. This can be done by sending an immediate expiration date along with every page served by a web site. This will have the effect that each time a user will try to access a cached page, his/her browser will realize that the cached page is outdated and that it needs to request a newer one from the web server. In short, the browser will never retrieve pages from cache.

One would expect that such monitoring may interfere with surfer behavior. That is, the delays caused by disabling caching are likely to frustrate surfers and cause them to leave a site earlier then they would if caching was enabled. Similarly, users who are sent cookies and do not want to register with a site are likely to leave it immediately. As a result, one would expect to see a sharp decrease in the average number of page requested by surfers in an environment where monitoring is taking place as compared to an environment that is free of monitoring.

To test this hypothesis, we set up a test web site with the ability to implement a 2x2 factorial design as shown in Figure 3. The site would dispense cookies to half of the visitors who are accessing it. It would also prevent caching during half of the visits. That is, when a visitor comes to our test site, he/she would have a fifty-fifty chance of being assigned to the cookie treatment. In this case he/she would be issued cookies, which he would keep for any subsequent visit. At the same time, each visit is assigned to a cache or no-cache condition. This second assignment is made at the visit level and not the visitor level. Hence, a surfer who visits the site repeatedly will make some visits in a cached environment, and some in a non-cached one. He will however see only one of the two cookie conditions.

==================

Insert Figure 3 about here

==================

We collected data for 1074 visits. 52% of these visits were made in a no-cache environment, while 67% were made in a cookie environment. Similarly to Drèze and Zufryden (1997) we used a Poisson regression approach to test whether cookies and/or cache disabling impact surfer behavior. Our dependent variable is the number of pages requested during a visit. Our independent variables are two treatment variables (Cookies and No-Cache) as well as a series of control variables that are not reported here for the sake of conciseness.

Our analysis shows that contrary to our prior belief, the use of cookies did not have a negative impact on the number of pages requested by visitors (see Table 7). Indeed, the regression coefficient for cookies is not significant. In contrast, the coefficient for the no-cache treatment is significant at the p=0.01 level. Moreover, this coefficient is positive. This means that more pages are requested in the no-cache environment. This makes sense given the nature of caching.

However, if one looks at the number of unique pages requested (eliminating duplicate pages requested within a visit), the coefficient for the no-cache situation becomes negative, at -0.21 (still significant at the p=0.01 level). This means that although more pages are requested, the breadth of the visits made by surfers in a no-cache environment is smaller than the visits made in a cached environment. This is an indication that visitors may resent the delays caused by the disabling of their caches.

TABLE 7: Poisson Regression of Number of Pages Requested as a Function of Cookie Issuance and Cache Disabling.

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]

Cookies	.08	.07	1.100	0.271	-.06	.22
No-Cache	.43	.05	8	0.000	.32	.54

CONCLUSION

This paper suggests that significant problems of advertising effectiveness measurement must be resolved before Internet advertising is “ready for prime time.” More specifically, our study has demonstrated that measurements based on current techniques are subject to significant errors and thus will tend to make Web-based advertising effectiveness difficult to compare to standard media. In order to insure comparability with standard media, we need to identify unique visitors rather than use current methods that rely on IP address identification. This can be done in a variety of ways: through password, a rather obtrusive method; cookies, an unobtrusive but controversial method (e.g., see Komando, 1997); or even better through unique serial numbers assigned to each individual copies of the browser programs (a feature said to be upcoming in future versions of Netscape and Internet Explorer).

Another problem we highlighted, with respect to current measurement methods, is that they do not record cached page requests. However, we have shown that it is quite important to account for page requests from cache to accurately account for the number of potential exposures to banner ads inserted on Web pages. Our investigation of the use of path recovery algorithms has shown that they are either ineffective, or unreliable (i.e., do more harm than good). Here, some mechanism which forces all page requests to be made from a Web site’s server, rather than retrieving pages from cache, would solve the problem of failing to record exposures on a log file because of caching. However, we have shown that this technique may detrimentally affect surfers’ behavior. Obviously, this solution increases the download times of web pages and may contribute to the impatience and irritation of surfers. The irritation that surfers may experience when faced with long download times has been suggested as a potential adverse effect which may decrease a surfer's interest in the contents of a Web site (e.g., Ducoffe, 1996). Perhaps a technique that allows for the partial caching of Web pages may offer a compromise to this potential problem. For example, on the one hand, components of a page that take relatively more time to download, such as graphics, video and sound files may be cached. On the other hand, the text portions of a page (which can be reloaded quickly) might have instantaneous expiration date so that the text of the page will always be requested from the Web site server and hence recorded. This method would provide a record of all potential exposures to Web pages whether or not they were partially cached. Furthermore, it is expected that such a method would not significantly increase the resulting download time, relative to complete page caching, as text may quickly be downloaded from a Web site server to a user's PC screen. Consequently, there should be minimum irritation and loss of patience on the part of Web surfers.

Alternatively, current problems could be solved if browser programs had a way to uniquely identify their users, and informed sites about the users' identities whenever they retrieved a page from cache. Once the current measurement problems are resolved, Internet advertising should indeed achieve its full potential as a significant component of a company's media mix.

REFERENCES

ADVERTISING AGE INTERACTIVE, Advertising Data Place, 1998. http://adage.com/interactive/articles/19980406/article6.html

Briggs, Rex and Nigel Hollis, "Advertising on the Web: Is There Response Before Click-Through?," *Journal of Advertising Research*, 37, 2, March/April (1997), 33-45.

Coffee, Steve and Horst Stipp, "The Interactions Between Computer and Television Usage," Journal of Advertising Research, March-April (1997), 61-67.

Drèze, Xavier, Kirthy Kalyanam, and Rex Briggs, "The Ecological Inference Problem in Internet Measurement: Leveraging Web Site Log Files to Uncover Population Demographics and Psychographics," 1998, working paper.

Drèze, Xavier and Fred Zufryden, "Testing Web Site Design and Promotional Content," *Journal of Advertising Research, Vol. 37, No. 2, 1997.*

Ducoffe, Robert H., "Advertising Value and Advertising on the Web," *Journal of Advertising Research*, 36, 5 (1996), 21-35.

ELECTRONIC MARKETPLACE AND ADVERTISING REPORT, January 1997.

Headen, Robert S., Jay E. Klompmaker, and Jesse E. Teel Jr., "Predicting Audience Exposure to Spot TV Advertising Schedules," *Journal of Marketing Research,* 14 (February), 1-9.

INTERNET ADVERTISING BUREAU, " Report on Online Advertising Revenue," April, 1998.

I/PRO CyberAtlas, 1996.

I/PRO-DOUBLECLICK, “Web In Perspective: Ad Response, What Makes People Click?,” *Advertising Research Foundation Interactive Media Symposium*, Monterey, CA, February 3-5, 1997.

JUPITER COMMUNICATIONS, “Ad Revenues Jump 83% in Second Quarter, According to Jupiter AdSpent Data,” Press release, September 3, 1996.

JUPITER COMMUNICATIONS, “Web Ad Revenues,” Press release, April 6, 1998.

Komando, Kim, “No Milk Needed With These Cookies,” *Los Angeles Times*, June 9, 1997, D5.

Metheringham, Richard A., “Measuring the Net Cumulative Coverage of a Print Campaign,” *Journal of Advertising Research*, 4 (1964), 22-28.

ONLINE MONEY NEWS, “Amazon.com and America OnLine Announce Multi-Million Dollar Advertising and Promotional Agreement,” Press release, July 8, 1997.

Pedrick, James H., and Fred S. Zufryden, “ Evaluating the Impact of Advertising Media Plans: A Model of Consumer Purchase Dynamics Using Single-Source Data,” *Marketing Science*, 10, 2, Spring (1991), 111-130.

Zufryden, Fred S., “A Model for Relating Advertising Media Exposures to Purchase Incidence Behavior Patterns,” *Management Science*, 33, 10, October (1987), 1253-1266.

Appendix A: Measurement Error for Three Web Sites

		Site 1			Site 2			Site 3	
	ID Based	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**
# of Visitors	709	576	-18.76	63832	52263	-18.12	25260	14006	-44.55
# of Visits	830	663	-20.12	90039	69290	-23.04	42997	29894	-30.47
Page/Visit	2.54	3.19	+25.19	8.89	11.67	+29.96	13.36	19.22	+43.82
Time Spent	60.98	98.46	+61.47	385	531	+37.92	396	536	+35.35
Repeat Visits	1.17	1.15	-1.68	1.41	1.33	-6.01	1.70	2.13	+25.39

		Site 4			Site 5		Overall
	ID Based	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**	**% Error**
# of Visitors	26099	9790	-62.49	14506	6870	-52.64	-39.31
# of Visits	36761	15183	-58.70	18437	10310	-44.08	-35.28
Page/Visit	17.38	42	+141.66	25.99	46.5	+78.91	+63.90
Time Spent	335	903	+169.55	274	529	+93.07	+79.47
Repeat Visits	1.41	1.55	+10.11	1.27	1.50	+18.08	+9.18

Appendix B: GRP Measurement Error for Three Advertising Alternatives

Site 1		Ad 1			Ad 2			Ad 3	
	ID Based	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**
Reach	0.25	0.34	+40.07	0.18	0.21	+20.16	0.30	0.43	+41.90
Frequency	2.31	2.03	-12.12	1.22	1.25	+2.44	2.21	1.92	-13.25
GRP	0.57	0.70	+23.09	0.22	0.27	+23.09	0.67	0.83	+23.09

Site 2		Ad 4			Ad 5		Overall
	ID Based	**IP Based**	**% Error**	**ID Based**	**IP Based**	**% Error**	**% Error**
Reach	0.0218	0.0245	+12.55	0.0094	0.0105	12.16	+25.37
Frequency	1.1541	1.2523	+8.52	1.2167	1.3249	+8.89	-1.10
GRP	0.0251	0.0307	+22.14	0.0114	0.0140	+22.14	+22.71

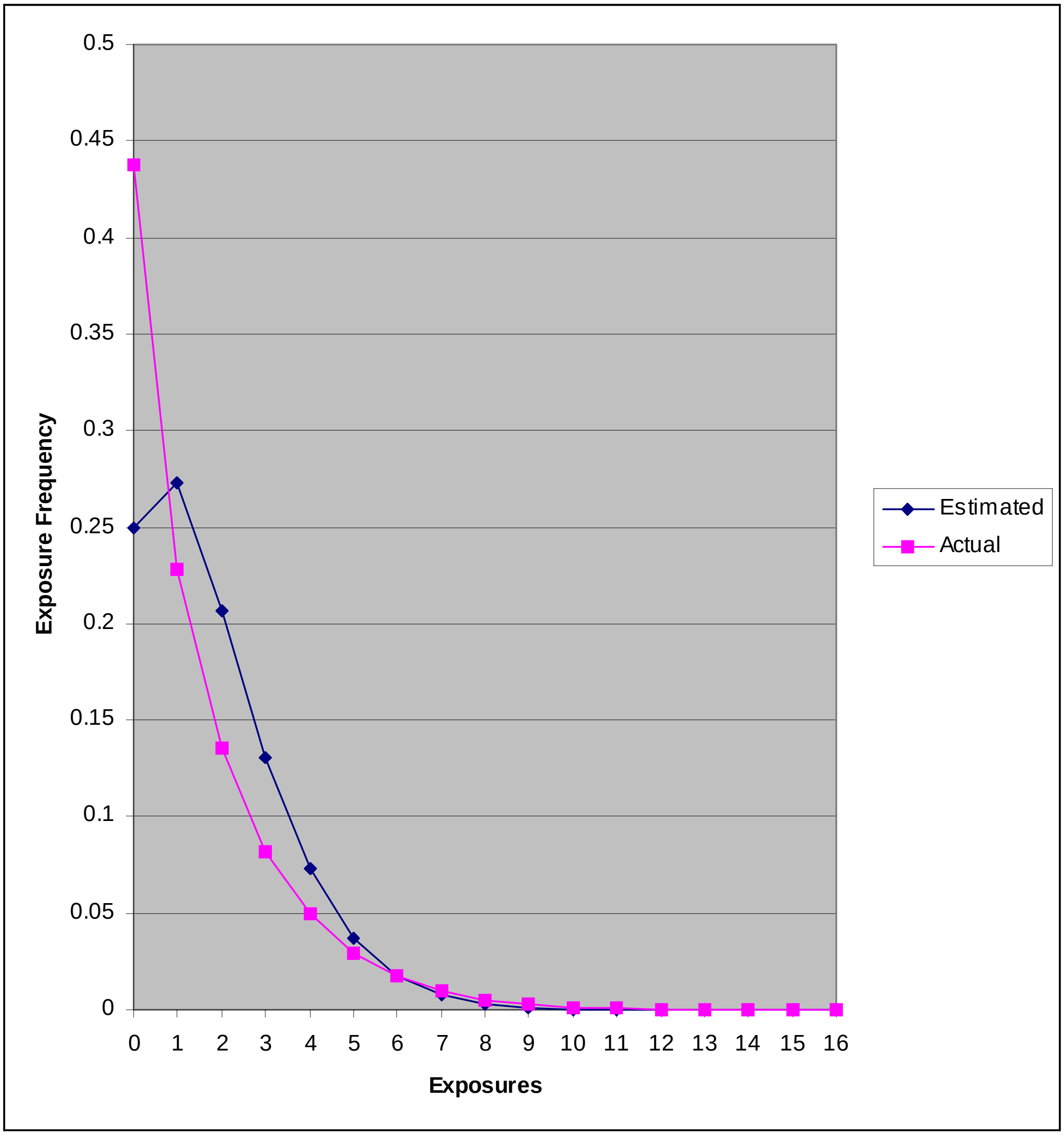
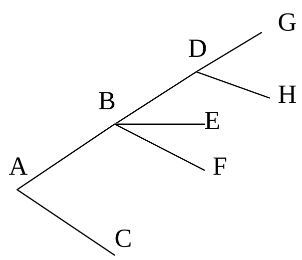


FIGURE 1: Contrast of BBD Exposure Distributions Based on Estimated and Actual Levels of Reach and GRPs.

Sample Site Visit



Visit: A B D G D[1] H D[1] B[1] E B F A[2] C
Log with Cache: A B D G H E F C
[1]Use the "Back" button
[2]Use the "Go" menu

FIGURE 2: Illustration of Cache Recovery Algorithms

Treatments:	Issue Immediate Expiration Date on Pages Served (i.e., No-Cache)	Do Not Issue Expiration Date on Pages Served (i.e., Cache)
Issue Cookies	n_{11}=388	n_{12}=359
Do Not Issue Cookies	n_{21}=169	n_{22}=158

* n_{ij} represent the sample sizes tested in each cell

FIGURE 3: 2x2 Experimental Design to Test to Study the Impact of Surfer Monitoring Procedures