VOL. 5, NO. 4, APRIL 2010 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

A THEME ON WEB MARKETING THROUGH WEB TRAFFIC

Nanhay Singh[1], R. K. Chauhan[2], Raghuraj Singh[1], Navin Kumar Tyagi[3], A. K. Solanki[4] and Deepak Kumar Swain[5]
[1]Harcourt Butler Technological Institute, Kanpur, India
[2]Kurukshetra University, Kuruklshetra, India
[3]Marathwada Institute of Technology, Bulandsahar, India
[4]Bundelkhand Institute of Engineering and Technology, Jhansi, India
[5]Apeejay Institute of Technology, Greater Noida, India
E-Mail: nsingh1973@gmail.com

ABSTRACT

In business, web traffic analysis plays a greater role and it is relatively a new market segment. With the development of World Wide Web (WWW), companies are beginning to get more sophisticated about how they use their web sites and want to know how to increase their return on investment (ROI). This paper reviews the business trends with web traffic analysis and presents marketing research about what is most important to companies and businessman in this segment and proposes to study the different parts that a global web-marketing approach uses and will also focus on the structure that can be deployed to support web marketing. Web-marketing is a multipart work, and it must always be linked to real world marketing. The days of building a web site to simply tell people about your business are over. Now, among plenty other features, you need user interaction by allowing them to access reports, communicate directly and purchase products. You need to be on time with your services and up-to-date with your information in relation to the marketplace. With true eBusiness; it is business requirements, rather than technical elegance, that drive the specification of the web site. Web sites become increasingly personalized to the constituents they are trying to support, namely, customers, investors, and employees. Web mining enables new understanding of user goals on the web. This understanding has broad applications and traditional mining techniques such as association rules have been used in business applications.Web marketing is the process of growing and promoting an organization using online media. Web marketing does not simply mean building a website or promoting a website. Somewhere behind that website is a real organization with real goals.

Keywords: web marketing, log analysis, traffic analysis, web mining.

1. INTRODUCTION

The Web has brought many unique benefits to marketing including low costs in distributing information and media to a global audience. The interactive nature of web media, both in terms of instant response, and in eliciting response at all, are both unique qualities of web marketing.

This paper deals with monitoring, collecting, analyzing, and modeling of World Wide Web (WWW) traffic [1] and client interactions to analyze the web marketing strategy [2]. The rapid growth of web usage has not been accompanied by an overall understanding of models of information resources and their deployment strategies. Consequently, the current web architecture often faces performance and reliability problems. Scalability, latency, bandwidth, and disconnected operations are some of the important issues that should be considered when attempting to adjust for the growth in Web usage. The WWW Consortium [3] launched an effort to design a new protocol that will be able to support future demands. Before doing that, however, we need to characterize current users' interactions with the WWW and understand how it is being used. The Web has become part of the fabric of our society, and accordingly we have an increasing need to understand the activities and goals of web users. We can improve nearly every aspect of the user experience on a web site by understanding the users' goal and traffic composition. Webmasters and content producers would like to gain an understanding of the people that are visiting their web sites in order to better tailor sites to user needs [4]. Marketers would like to know the users' interests in order to have better sale promotions and advertisement placements [5]. News sites would like to produce and present materials that are highly relevant to their visitors. No matter how we measure traffic on web, the growth of the web is phenomenal. The number of web users [6] is now measured in the tens of millions while the number of web sites is now measured in the millions. In fact, the web is growing so quickly that it is useless to quote growth statistics, as no one can consistently quantify or agree on the actual figures! Regardless of the actual numbers, it is clear that a lot of time, attention, and money have been spent on the web by companies wanting to get involved in cyberspace [7]. Very few of them have much of a feel for their payback on this investment. Much of that has been due to the incredible hype and fast growth surrounding this technology, combined with the low cost of experimentation. All one really has to do is develop a web site and revenues will shoot through the roof. The web is fundamentally different that most other marketing venues.

In traditional marketing [8], it's relatively easy to construct a profile of your target audience in terms of demographics and psychographics [9]. With the web, however, you don't know much at all about the visitors to your web site - as most visitors are anonymous. They come and go from your site without much of a trace. The more sophisticated marketers try to collect demographic

information anyway, so that it can be used to target future content and messages. This is typically done by offering visitors something of value in exchange for filling out a form. This needs to better understand visitors and what they do on a web site is behind the increased demand for web traffic analysis. This approach provides behavioral data about visitors, including the types of companies that visit your site, where they came from, and what they do when they get there. As the number of internet users and sites increase, as the number of page requests increases exponentially. Web traffic analysis can help to draw conclusions from this mind-boggling volume of behavioral data. The ultimate goal is to combine this behavioral information with traditional demographic and psychographic information. This allows measurement of what people say, how they feel, and most importantly, how they actually respond (which are not always consistent or even logical). This information is the foundation of personalized one-to-one marketing techniques [7], allowing a business to target specific audiences with customized products and services that directly solve their problems. We call this web-mining [10] (i.e. using web behavioral information as an integral part of your one-to-one marketing).

2. WHY WEB TRAFFIC ANALYSIS

Web traffic analysis is the process of measuring extent and character of activity of users on a website, interpreting the measurements, and applying the conclusions. Web traffic analysis is crucial for any organization that has a web site. For example, commercial corporations or the private sector want to increase revenue and profitability, while government agencies, universities and colleges or the public sector want to provide services and information. Even though theses two sectors have different goals they do have one goal in common. They must be able to provide the highest level of service to their users. This can be achieved by:

- Providing the most relevant information to the people visiting the web site.
- Improving communication via the intranet and extranet [11].
- Optimizing the buying process to maximize revenue.
- Promoting the information that is the most popular.
- Creating a convenient experience by organizing information more effectively on the site.
- Improving user satisfaction by reducing the number of clicks it takes to access information.

Web traffic analysis is the study of web user usage patterns. This will allow an organization to determine:

- Where are my web users coming from?
- What path do they take before finding the information they want? What path do they take before leaving?
- Which pages are the most popular?
- Which pages are the least popular?
- How many clicks did it take to find the information?

But before these questions can be answered a process and/or application is needed to capture, store, manage, analyze, and report web usage data.

3. BACKGROUND

The most easily measured benefit from eBusiness [12] is the cost savings associated with allowing customers to serve themselves. Communicating purchase requirements, checking order status, selling on-line, and providing timely content are all good examples of this. In addition, eBusiness can increase profits through increased revenue generation. Common examples include direct product sales and advertising revenue. Customers want to increase reporting capabilities to support more complex analysis needs - typically to support return on investment (ROI) [7] calculations. This analysis involves gathering data about the market and adding it to the analysis. The type of data that is added includes advertising groups and names, filters, and virtual server information. This data is used to parse and filter log data [13], as well as report advanced relationships of grouped and named data. Product managers and marketers are typical candidates for tools in this segment. They are typically managing sites at the customer support and on-line transaction stages of development. Rather than look at the web as its own island of information, customers want to integrate with other databases in the corporation such as customer databases. Resulting reports not only use advanced relationships between log data, grouped and named log data, but also draw from external databases and spreadsheets. The highest level of customers gather data directly from the wire, eliminating log files as the primary source of information. Many web miners adapt their offers based on traffic and, more importantly, create new products that match the results of their analysis. For these people, web traffic analysis is gold! Business owners and managers are typical candidates for these tools. They typically have sites at the personalization and community stages of development. The web traffic analysis market barely existed a short five years ago. Today it is a multi-million dollar segment of the software industry. Clearly, this is being driven by the growth of the World Wide Web (WWW) and the desire to know as much as possible about visitors, through self-identification, registration, and web logs. Hence in the field of web marketing the structural web marketing [8] have also greater role. It simply looks like the following (Figure-1):

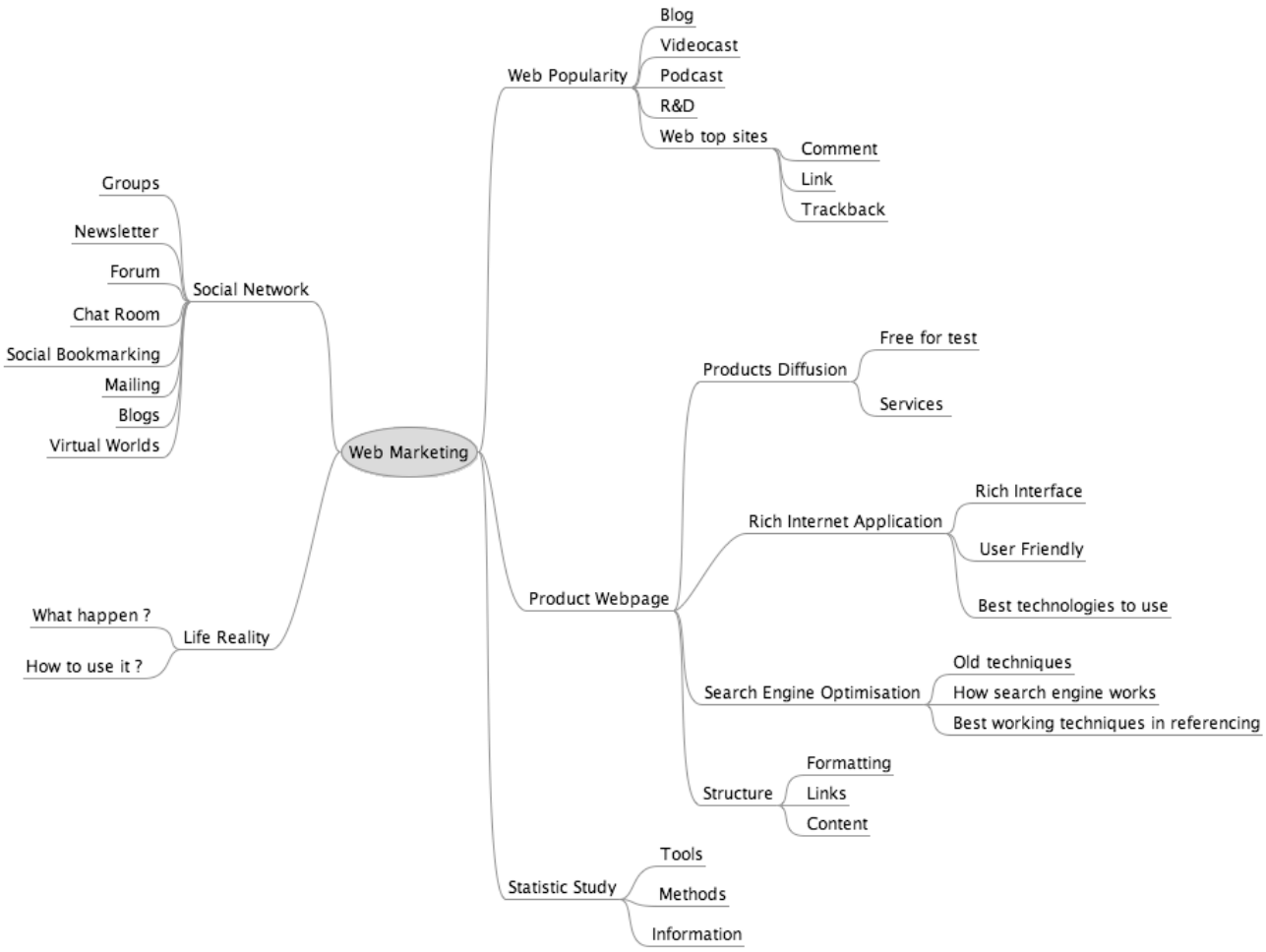

Figure-1. Structural web marketing.

4. METHODOLOGY

There are four techniques to analysis the information over the web:

a) Log file analysis
b) FTP log files to vendor
c) Client side scripting
d) TCP/IP packet sniffing

Usually Log File Analysis based software is of highly capable system, inexpensive, easy to maintain and limited capability. So what exactly is a web log, and where does this data come from? When you browse the web, you are merely making an internet connection with a remote computer (called a server) and asking for one or more files to be sent to you via TCP/IP (Transmission Control Protocol/Internet Protocol) [14]. This protocol spits the file data into little pieces called packets and sends them to your computer via the Internet. These packets are then reassembled by your computer and displayed via your browser. Behind every web site you visit is a web server, whose purpose in life is to respond to you by locating and sending the requested files to your computer. After each request, the web server logs the results of the exchange in a log file. A typical log file contains information about which computer made the request, for which file, and on which date. Additional information is sometimes recorded as well including the browser type, the IP address, error codes, the referring site, and the destination site. Log file analysis tools simply take this information and try to make sense of it so that intelligent conclusions can be drawn. Simple things like how many total files were requested can be easily calculated and reported. By looking for multiple requests from the same computer during the same timeframe, more complex things can be calculated, like the number of total visitors and visits that were made to a site. By adding other information, like advertising information, ad impressions and click through rates can also be calculated.

5. AN EXAMPLE

Log file (Combined log format)
Limestone.uoregon.edu andred - [19/Jun/1999:00:49:41 -0500] "GET/service/contracts.gif HTTP/1.0"200 1341 "http://www.netgen.com/"Mozilla/2.0 (compatible; MSIE 4.0; AOL 4.0; Windows95)"

Description of above example:

- Hostname or IP address - limestone.uoregon.edu
- Registered user name (usually blank) -…….

- Date and time of request -[19/Jun/1999:00:49:41 -0500]
- Object requested -"GET/service/contracts.gif HTTP/1.0"
- Status code -200
- Bytes transferred -1341
- Referral information -"http://www.netgen.com/"
- Browser information - Mozilla/2.0(compatible; MSIE 4.0; AOL 4.0; Windows95)" (Courtesy netGenesis Corp)

There are many tools available in the market for web traffic analysis. With the help of these tools many type of reports can be generated and hence can be used to analyze the data for the purpose of research work. The lists of these tools are shown in the Annexure A in Table-1.

6. IMPLICATIONS AND FUTURE WORK

Accesses to web proxies [15] are highly correlated from day to day and from one week to another. A key source for this correlation is the daily and weekly schedule of web users. Users who share the same time zone should have similar schedules and hence part of their accesses should reflect it. This result is important for commercial proxies since they will act similar to a local broadcasting station or local cable company that distributes information and other kinds of media. Proxies will be installed to support a group of users in a certain locality. Peak and minimum usage hours can be identified for upgrades, price commercials, and other time dependent jobs. The fact that web traffic has a significant deterministic and cyclic component is very important for internet service providers (ISP). This fact can be used for predicting peak hours and hence to schedule software upgrades or maintenance. It also can help in the efforts to smooth and distribute the traffic to the service provider proxy by advertising the low activity times and having lower pricing rates then to encourage and attract more users. Internet network protocol designers can use this information to support pre-fetching and filtering of data based on the daily and hourly level of activity and locality of reference of a group of clients connected to a certain proxy. This result is also important when characterizing accesses to busy web servers. The traffic to such servers can be split based on the time zones of the clients accessing the server. Then the characteristics of sub-groups can be modeled using the suggested approach. The discovered periodic correlation indicates that queuing and statistical models are not the appropriate techniques to model such systems. Simulations need more accurate models that represent the characteristics of the web traffic. There is a new modeling approach that captures the real characteristics of the web traffic by using a combination of Fourier analysis and the traditional techniques of statistical analysis [1]. The new approach characterizes the random and the deterministic parts of the data. The generated data displays long-range dependent and self-similar characteristics comparable to the ones observed in the original data.

7. CHARACTERIZING USERS' SEARCHES AND SESSIONS

In this session we analyze accesses from groups of clients to several Web Information Retrieval Systems (IRS), [16] and in particular study their queries. To help users locate information on the web, special search tools and cataloging systems were developed. We call these tools Web Information Retrieval Systems (IRS). They have evolved rapidly since 1994, and have been used by millions who thereby had their first experience with search engines. Borgman *et al.* [1] observed that the end users who now dominate searching are using systems with exploratory interfaces, under less time pressure, and have less clear retrieval goals than do skilled search intermediaries. For the characterization of WWW traffic so as to support modeling, planning and prediction, and to help scale up the web, it requires the study of retrieval-related traffic [1]. In particular it requires the study of sessions, queries, and browsing activities using log files; the log files come from diverse situations. It characterizes accesses to different web IRS to find out the most popular systems. Next, the study of client systems to see how much web traffic is due to multi-user systems and if such systems have similar access patterns to those of individual users. Then it characterizes queries by looking at their complexity bandwidth than other sessions. There are many algorithms to extract client sessions. Second, it analyzes the sessions and show that those with more search steps are responsible for fewer transferred bytes. It defines user sessions in terms of browsing, searching, and next step activities. Then it looks at the most popular patterns in the identified sessions. Using regression, it discovers a correlation between amount of searching in a session and the bandwidth requirement with respect to numbers of terms and operators. Finally, it demonstrates that sessions with searching tend to consume fewer networks.

8. FINDING INFORMATION OVER THE WEB

Currently users can look for information over the web using three methods:

8.1 Browsing

Users can follow links from page to page, reading or looking at what they find interesting. However, even those with a focused information need may get distracted during the browsing process and end up reading or looking at something completely irrelevant to their initial interest. On the other hand, browsing can be particularly efficient when users are looking for information within a specific collection that is limited in size.

8.2 Using web catalogues

Like with Yahoo, users with a general topic can apply a classification system provided to narrow their subject until they end what they are looking for. For example, if we are interested in HTML and want to read about it, we can take the following path to get to the required information: from the main directory in the Yahoo home page to Computers and Internet, World Wide

Web, and finally HTML. Following these three steps will lead to several further choices about HTML including editing tools, standards, manuals, etc.

8.3 Using web IRS

When looking for something more specific, users often turn to search engines such as Lycos or Google. For example, we may want to read a certain article and know the author of the article but not know the journal or date. We can use the author's name and a key word from the article to locate it. While the web is used in other ways, these three methods cover a high volume of web traffic, and so are worth more careful characterization.

9. RESULTS AND OBSERVATIONS

For users who run on-line advertising, can track and compare results. This provides verifiable advertising information such as click-through rates, ad impressions and ratios. It helps web marketers understand the quality of their ad banners by reporting visitor's path, from click through to exit page.

9.1 Keyword reporting [7]

It tells users what visitors were searching for in search engines like Yahoo to find your web site. This is an effective feature for companies wishing to improve their promotion and presence on the popular search engines.

9.2 Query string parsing [7]

It allows web marketers to understand what type of information their visitors are searching for in databases on their own site. It is popular to provide dynamic content where pages are served by a database. Query string parsing is valuable information for repositioning content, developing marketing events, etc. It is essential for database driven sites that use Microsoft ASP pages or Allaire Cold Fusion.

Sophisticated web analysis users want to be able to analyze data by market segments, rather than simply as a whole. Cross-tabbing provides the ability to cross reference and compare one or more data elements - which is important if you want to compare visitors by country against which operating systems are most popular in each for example.

Many users want higher level reports, using groups, so that related information can be combined. This puts information into more meaningful and readable top-level groups. Typical examples include showing all visitors from AOL, CompuServe, Prodigy and MSN in a single group called online services. The web marketing analysis through web traffic provides the reports which are pre-defined reports that already do most of customers want. In general, customers don't want to spend lots of time defining new reports and dealing with reporting layout. This analysis of web marketing through web traffic provides users to easily build complex reports from a wide variety of web traffic analysis calculations.

9.3 Six types of users

User Type 1: Trackers - 37%

- Tracking past shipments
- Characterized by low duration.

User Type 2: Reservers - 3%

- Complete online reservations.
- Low duration per page view.

User Type 3: Uncommitted - 10%

- Characterized by long duration.
- Fail to complete transaction.

User Type 4: Info Gatherers - 4%

- Concentrated in information areas.
- Rarely reach transaction areas.

User Type 5: Single-clickers - 32%

- Visit homepage only.
- Not qualified customers or prospects.

User Type 6: Wanderers - 15%

- Very few, very random pages.
- Few hits, but long duration per page view.

(Courtesy netGenesis)

On the basis of the above six type of users we can assume that how the web traffic leads to the web marketing analysis and the behavior of users.

Figure-2 illustrates the behaviors of users when they are purchasing something on the web. The total number of user reduces step by steps.

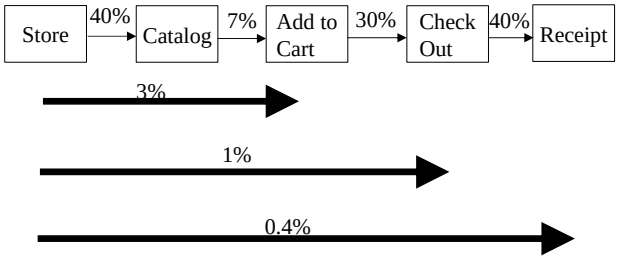

Figure-2. Percentage of users actually purchasing.

Also on analysis of different commercial sites different data found and hence the reports are made. The examples of some reports are shown in the Annexure-A. Figure-3 simply tells about the most visited pages, Figure-4 describes top hostnames with number of page views, Figure-5 illustrates the Clickstream example by calculating number of visits of different pages of a web site.

10. CONCLUSIONS

This study provides the efficient way to analyze the marketing behavior of customer on the web**.** As analysts, we are deeply involved in making the web more accessible to users. We need to know what users are doing in order to better optimize the web sites. Recent research seeks to understand the composition of user traffic using web mining techniques on server logs. The commonality is to first build up user profiles based on the user visitation paths, and then apply clustering techniques to these user profiles. This paper summarizes the access of several websites for a specific time period and hence analyses the customer behaviors. First, there will be a framework in which any of the available data features can be used in the clustering of user sessions. This framework enables robust combinations of content, linkage structure, and usage to discover user needs. Second, there should be results from a systematic evaluation of different clustering schemes by conducting a user study. By knowing what the tasks were and how they should be grouped in advance, post-hoc analysis [17] of the effectiveness of different clustering schemes were done. By counting the number of correct categorizations, certain combinations of clustering schemes enabled to obtain accuracies of up to 99%. While is not necessarily expected this same level of accuracy on real world web logs, which are much noisier by nature, this is still quite encouraging to analysts hoping to make sense of user actions on the web. This knowledge leads to built Lumber Jack [17], a prototype automated analysis system that couples user session clustering with more traditional statistical analyses of web traffic. This combination seeks to not only identify significant user information goals, but also to understand how users follow these goals as they surf the site. The goal is to create a completely automatic system that quickly and accurately identifies a site's dominant usage patterns. In the quest to understand the chain of user interactions on the web, analysts need tools for getting quick and accurate pictures of web usage. The work presented here suggests that in many cases the structure of user activity may be inferred automatically.

11. RELATED WORK

Most of the research that uses proxy traces aims to reduce network latency, enhance response time and conserve network bandwidth. Examples of such studies are systems and simulations using server [AW96b, BC94, KMR95a, AKMS95] and proxy traffic [ASA+95, WAS+96, Smi94, O'C95]. In those studies the authors have shown that caching will reduce network traffic. There have been several studies to characterize client workloads [CB96, CBC95, Cun97] and server workloads [AW96b, BC96]. There are, however, fewer studies to fully characterize proxy traffic. This is due to the difficulty of collecting proxy log files from different sources due to the sensitivity of such logs. One of the earliest studies to characterize proxy traffic was done by Glassman [Gla94]. Characterization included parameters such as document popularity, cache misses, cache hits, and rate of change for Web pages. Sedayao [Sed94] studied and characterized the distribution of several parameters. Gwertzman and Seltzer [GS96] have used several server and proxy traffic sources to characterize MIME types and the average life span of a Web document. A study that characterizes accesses of dial-in users with modem connections was done by Gribble and Brewer [GB97]. They have focused on inter-arrival time and discovered that the traffic is periodic on large time scales (hourly, daily and weekly). Arlitt and Williamson [AW96b] used six different log files to characterize accesses to WWW servers. From these logs the authors identified ten different invariants for web server workloads. The invariants are important since they aim to represent universal truths for all Internet Web servers. The invariants in the study were used to identify two strategies for cache design and to determine the bounds on performance improvement due to each strategy. A study to test if those invariants hold true over time and across location is necessary. In [CB96] and [CBC95] the data were collected from a group of clients accessing the Web. Cunha *et al*., [CBC95] showed that many characteristics of the WWW can be modeled using power-law distributions. Crovella and Bestavros [CB96] showed that the Web traffic has characteristics that are consistent with self-similarity. They traced the reasons for Web self-similarity to the basic characteristics of information organization and retrieval. Finally, Pitkow [Pit98] comprehensively summarizes work done in the field of workload characterization

REFERENCES

[1] Ghaleb Abdulla, Edward A. Fox, Chair. 1998. Analysis and Modeling of World Wide Web Traffic. Computer Science and Applications, Virginia Polytechnic Institute and State University.

[2] Traffic Analysis of a Web Proxy Caching Hierarchy Anirban Mahanti, Carey Williamson**,** and Derek Eager, University of Saskatchewan (published by IEE).

[3] J. Gettys. 1995. HTTP problem statement. http://www.w3.org/-Protocols/HTTP-G/951005 Problem.html.

[4] Yan T.W., Jacobsen M., Garcia-Molina H. and Dayal U. 1996. From User Access Patterns to Dynamic Hypertext Linking. Computer Networks. 28(7-11): 1007-1014.

[5] Barrett R., Maglio P.P. and Kellem D.C. 1997. How to personalize the Web, In: Proc. of the ACM Conference on Human Factors in Computing Systems, CHI '97 (Atlanta GA). March. pp. 75-82.

[6] Introducing Traffic Analysis Attacks, Defences and Public Policy Issues. (Invited Talk) George Danezis

K.U. Leuven, ESAT/COSIC, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium.

[7] Mike McClure, V.P. Marketwave. Web Traffic Analysis Software- An Industry. White paper: 206-682-6801. http://www.marketwave.com

[8] Ahmet Gyger. 2007. Licence Thesis Structural Web-Marketing. Under the direction of Michel Deriaz Supervisor: Professor Dimitri Konstantas Mars.

[9] Kalakota R. and WHINSTON A. B.: \Do or Die: market Segmentation and Product Positioning on the Internet. http://cism.bus.utexas.edu/res/articles/segmentation.html.

[10] Department of Computer Science 200 Union Street SE, 4-192, EE/CSC Building University of Minnesota, Minneapolis, MN 55455, USA_ srivasta_, desikan_, kumar_@cs.umn.edu, http://www.ieee.org.ar.

[11] M. Baentsch *et al*. 1997. World-Wide Web Caching-The Application level View of the Internet. IEEE Commun. Mag. 35(6).

[12] Anderson Richard and Jared Braiterman. 2001. Strategies to Make E-Business More Customer-Centered.

[13] W. Heagy. 1998. WWW server log data visualization and evaluation. Master's Thesis. Computer Sci. Dept., Virginia Tech, Blacksburg, VA 24061-0106.

[14] W. Richard Stevens. 1994. TCP/IP Illustrated, Volume I: The Protocols. Addison Wesley, CA.

[15] A. Luotonen and K. Altis. 1994. World-Wide Web proxies. Computer Networks and ISDN Systems. 27(2). http://www1.cern.ch/PapersWWW94/luotonen.ps.

[16] Christine L. Borgman Sandra, G. Hirsh and John Hiller. 1996. Rethinking online monitoring methods for information retrieval systems. Journal of the American Society for Information Science. 47(7): 568-583.

[17] Lumber Jack. 2002. Intelligent Discovery and Analysis of Web User Traffic Composition. (Ed) H. Chi, Adam Rosien, Jeffrey Heer. PARC (Palo Alto Research Center), Coyote Hill Road, Palo Alto, CA. http://www2.parc.com/istl/groups/uir/publications/items/UIR-2002-10-Chi-WebKDD-LumberJack.pdf.

Annexure-A

Table-1. Commercial web traffic analysis tools.

Tool	Method	Comments
Net genesis www.netgenesis.com	Log file analysis	Highly capable system
Wusage www.boutel.com/wusage	Log file analysis	Inexpensive, easy to maintain ,limited capability
Accure www.accure.com	Log file analysis	Troubles with earlier versions
Net acumen www.netacumen.com	FTP log files to vendor	Privacy issues. website for reports
Hit box www.hitbox.com	Client side scripting	Good for simple sites
Aria www.macromedia.com	TCP/IP packet sniffing	Does not use log files

Top Pages Example

HTTP Resource	# of Page Views	%of Total	Cum %
1 /	2,136,650	38.5	38.5
2/bigadmin/downloads/	228,679	4.1	42.7
3/MySun/	198,430	3.6	46.2
4/bigadmin/docs/	131,694	2.4	48.6
5/search/index.cgi/	103,248	1.9	50.5
6/staroffice/	65,347	1.2	51.6
7/products-n-solutions	65,038	1.2	52.8
8/corp_emp/scripts/showjob.cgi/	63,601	1.1	54.
9/products/staroffice/get.cgi/	60,260	1.1	55.0
10/forte/ffj/overview.html/	58,103	1.0	56.1
11Other	2,434,788	43.9	100

(Altered data.)

Figure-3. Top pages example.

Drill Down, Top Hostnames Example

Top Hostnames for /corp_emp/scripts/showjob.cgi, for time period in previous report:

Hostname	# Page Views	%of Total	Cum %
1 serv3.hwka.com	17,161	27.0	27.0
2 209.67.186.119	5,998	9.4	36.4
3 216.34.97.92	5,736	9.0	45.4
4 ip22.digibahn.net	1,103	1.6	47.0
5 areil.sun.com	501	0.8	47.8
6 mailgate.cwhkt.com	363	0.6	48.4
7 pix89.pgexch.com	249	0.4	52.2
other (5152)	30,115	47.2	100

(Altered data.)

Figure-4. Drill down, top hostnames example.

Clickstream Example

	Number of Visits		% of Tota
First page	**11507**		**100.00%**
Second page	**9096**		**79.00%**
Third page		**7000**	**61.00%**
Third page		**1500**	**13.00%**
Third page		**500**	**4.30%**
Third page		**96**	**0.80%**
Second page	**1214**		**10.60%**
Third page		**577**	**5.00%**
Third page		**394**	**3.40%**
Third page		**134**	**1.20%**
Third page		**109**	**0.90%**
Second page	**1137**		**9.90%**
Third page		**800**	**7.00%**
Third page		**200**	**1.70%**
Third page		**100**	**0.90%**
Third page		**37**	**0.30%**
Second page	**20**		**0.20%**
Third page		**10**	**0.10%**
Third page		**5**	**0.05%**
Third page		**3**	**0.03%**

Figure-5. Clickstream example.

(Courtesy netGenesis)