# Journal of Internet Banking and Commerce

*An open access Internet journal (http://www.arraydev.com/commerce/jibc/)*

*Journal of Internet Banking and Commerce, April 2011, vol. 16, no. 1 (http://www.arraydev.com/commerce/jibc/)*


## Identifying Internet Abuse by Analyzing User Behavior on the Internet


**Ahmad Mashhour, PhD**
Associate Professor, MIS Department, Yarmouk University, Irbid, Jordan
*Postal Address:* **Ahmad Mashhour, MIS Dept, Yarmouk University, Irbid, Jordan, postal code 21163**
*Author's Personal/Organizational Website:* ***http://www.yu.edu.jo***
*Email:* ***mashhour_ahmad@yahoo.com*** **(please use to correspond with the authors)**
Dr. Mashhour is currently a faculty member in the Management Information System Department, College of IT, Yarmouk University, Jordan. His current research interests are on E-learning, E-business, and simulation modeling using database systems.

**Abdulla Al-Saad**
Mr. Al-Saad is an ex-student in (post graduate study) at Yarmouk University, Department of Management Information Systems, Jordan. Currently employed at the Jordan Government Sector. His research interest is E-commerce applications.
Email: **lsdu51@moj.gov.jo**

**Zakaria Saleh,' PhD**
Dr. Saleh is a Faculty of Information Technology College, Management Information Systems Department (Associate Professor). His current research interests are on wireless networks, Internet Banking, and systems security.
*Postal Address:* **Zakaria Saleh, MIS Dept, Yarmouk University, Irbid, Jordan, postal code 21163**
*Author's Personal/Organizational Website:* ***http://www.yu.edu.jo***
*Email:* ***drzaatreh@netscape.net***

## Abstract

Advances in the banking sector have ushered in an era of multi-product and multi-services being delivered using multiple yet integrated channels. The use of information and communication technology (ICT) is on the increase and encompasses nearly the entire gamut of banking operations. Rising competition and customer expectations have compelled top management to implement, and continuously upgrade, agile and scalable ICT practices and solutions. The enormity and range of banking services combined with the complexity of integrated and ICT-enabled delivery mechanisms require comprehensive partnerships to be forged between banks and providers of ICT solutions, especially with regard to Business Continuity Management (BCM). This paper presents the preliminary findings of a research study to identify the essential ingredients of successful BCM implementation based on experiences of banks in India.

Keywords: **Firewall Log file; Internet abuse; Internet control; Log files analysis; User behavior**



---

### 1. Introduction

With the expanded use of Information Technologies (IT), employees in any organization can quickly and easily visit Web sites from their work place. The Internet gives these employees the ability to access useful resources and retrieve vast amount of helpful information in performing their jobs. However, the Internet can be misused, which decreases productivity, negatively affects customer service, drains network resources, and in some cases exposes organization to legal liability. A study conducted by American Management Association (Management Association, 2008; Flynn, 2005) reveals that 64% of employees use the Internet for personal interest during working hours. On the other hand, organizations management is usually proactive to control Internet usage, and may follow different strategies to control Internet by using monitoring and access blocking software, implementing Internet acceptable use policy, and educating employees about utilizing Internet in work place to communicate policies and aid in the prevention of the Internet abuse Young & Case, 2003; Wen & Gershuny, 2007; Snapshot Spy, 2010).

This study will provide management with a tool that will enable them to identify Internet abuse patterns by its employees by analyzing the log file, and then determine what use was appropriate and what was inappropriate when any employee accesses the Internet. The decision concerning appropriateness /in appropriateness of a Web site is dependent on the results of analysis of many factors including type of Web site and its relation to the job role of the user, volume of downloads, and day period of downloads.

### 2. Significance of the study

Researches on Internet usage by employees indicate that employees do not like to be monitored (Urbaczewski, 2002), but monitoring for the purpose of giving feedback was more acceptable to employees than monitoring to control (Panko & Beh, 2002). Log file analysis is a systematic

approach to examine and interpret the content of behavioral data. Its goal is to assist in finding patterns in the behavior of people as they interact with a computer application (Facca & Lanzi, 2005; Akman & Mishra, 2010).

Many studies use log file analysis to identify user behavior in Web usage, and most of these study analyze Web log files to improve Website structure and its services by making it more personalized (Gaskin, 1998). A unique study by Johnson and Chalmers (2007) use firewall log file to identify Internet misuse by classifying Web sites as appropriate / inappropriate without considering the job category of the employees. Johnson and Chalmers (2007) stated that using firewall log file to provide feedback and identify employee Internet abuse is limited and can only provide information about users' requests, which limits the ability to generate report about employees' Internet usage. For example, if the log record shows that a certain user visited an online auction site, this doesn't mean the user was abusing, perhaps this employee works in the purchasing department, and a significant part of his/her job is to find best prices for his company; this not considered in the above mentioned research. Another example, if a user visited a pornography Web site, if it showed up as a single, one-time entry, this may not necessarily signal abuse, rather it may be either a mistake (someone misspelled a Web address) or it may be unwelcome pop-up window. In either case the employee immediately closed the window and did not return to the offensive site. None of the earlier studies used the log file for producing an "abuse report" on employees Internet access activity, or determine if there was an abuse, by simply using the log file.

## 3. Literature Review

Using Internet and e-mails for non-work related, or for personal use is called Internet abuse, cyber loafing, Non-Work Related Computing (NWRC), and Cyber Slacking. In the modern office, the Internet links the organization to the outside world, making it easier and more cost effective for employers to coordinate global activities among customers and suppliers. This technology allows co-employees to work in different geographical locations (Gupta, 2004).This is because distance-separated team members can be monitored and connected by communication technologies that allow supervision even at a distance. The connection to local/wide area network could extend the sharing of files, documents and other information beyond the domains of the company. Also, the vast array of information that the Internet contains can be useful for many types of work and serve as catalyst for innovation and new ways of thinking. It allows organizations to work with customers across the globe at a minimal cost. It can also promote improvement in employee efficiency, productivity, and extend training and professional development to employees without high travel expenses and much time taken away from work (Gee-Woo & Swee, 2009).

Internet abuse results in lost wages through decreased productivity, furthermore, employees' personal use of company technology can flood computing resources, which in turn clogs bandwidth and degrades systems. Internet abuse can also put the organization at risk if the employee engages in illegal activities online (e.g., downloading software) or creates a harassing environment through viewing or sending offensive material (Lee, Lee, & Kim, 2004). Estimates are that between 20% and 30% of companies have fired an employee for Internet abuse including accessing pornographic sites, online gambling, and online shopping (Carl & Young, 2004).

There are several studies that discuss efficiency of management strategies and practice in work place. Christine, Kohut, & Booth, (2009) demonstrate how electronic use policy design can affect employee fairness perceptions and cyberloafing. Employees will be less likely to cyberloaf if the

company policy includes periodic monitoring, and conversely employers avoid implementing policies with disciplinary procedures allowing for managerial discretion or lacking appeal processes. Most firms, according to the American Management Association (AMA) utilize some form of an Internet use policy (AMA, 2008). Internet use policy are written documents which deal with employees rights and responsibilities regarding information technology and are designed to deter abuse of a company by elaborating what actions are acceptable to the organization. It is also intended to protect company and its employees from illegal and unethical behaviors streaming from abuse of Information Technology (Gaskin, 1998).Yong & Case (2003) consider Internet policy as a strategy to prevent abuse and found only 48% of organizations implemented Internet policy in work place, which is perceived effective strategy to prevent Internet abuse. This is because policy is not based on theory, and not communicated properly with employees (Christine, Kohut, & Booth, 2009) but rather on anecdotal advice (Young, 2010). Several studies show efficiency of Internet policy depends on many factors including organizational size and seriousness of discipline part of policy (Yong & Case, 2002). Gee-Woo et al., (2007), examined the role of task characteristics and organization culture in non-Work Related Computing (NWRC); the study reveals that ineffectiveness of NWRC control mechanisms occur under high degree of nonroutineness tasks and a fit between discipline systems and organization culture leads to higher employee satisfaction with NWRC management which subsequently leads to lower time spent on NWRC.

A growing trend suggests that a number of corporations rely upon Internet use policies to cut recreational use of the Internet during work hours and to mitigate legal liability regarding such misuse (Young & Case, 2003). Firewall log file usually contains behavioral data about user when navigate Web sites to provide feedback about Internet usage by employees in an organization. Log file analysis is the systematic approach to examining and interpreting the content of behavioral data. Its goal is to assist in finding patterns in the behavior of people as they interact with a computer application or the Internet (Facca & Lanzi, 2005).

Many studies use log file and analyze it to identifying user behavior in Web usage, most of these studies analyze Web log file to improve Web site structure and improve its service by making it more personalized (Gaskin, 1998). In literature, unique study by Johnson and Chalmers (2007) use log file to identify Internet abuse based on only categorizing entry URLs into appropriate and inappropriate categories by using commercial software, and then generate report for each category to show abuse in using the Internet. This study is only use URLs category to identifying Internet abuse, but there is another terms of abuse such as, time spend on using Internet, download size, and file type (music, videos, etc.).

## 4. Research Design

Log files are limited to only the data which can be determined by the system. They do not include data about other programs running on the employee's computer, and its relation to employee's job function, gender, age, etc., such data would be helpful in gaining an understanding of different aspect of the current e- business.

This research will introduce a system that will analyze Internet abuse pattern by combining data from log file with employee information and company policy, and add value to the monitoring tools by determining the employee abuse, not only by using the row data in the log file, but also by analyzing data in the log file and linking it to the employees job functions as well as organization's policies. All information is stored in a structured database and analyzed using statistical software.

**4.1 Research Methodology**

The implemented system will work based on the following criteria:

- Classify employees into groups with common job descriptions, job duties and position;
- Identify criteria for good use of each group in terms of time spent on Internet and download size and according to management policies;
- Use the log file as a core source for the evaluation;
- Data were stored in a database management system and analyzed utilizing statistical measures.
- Categorize visited Websites i.e. *Uniform Resource Locator* **(***URL***)** as appropriate and inappropriate;
- Use MS ACESS, or any other database management systems (DBMS), and then map the file records into appropriate fields in the database, and associate file records with classified employee groups within the database.
- Develop an application using ASP.Net that is capable of conducting the analysis.

The main purpose for transferring the log files into the database is because the database systems provide the ability to derive more information and get more adequate, easier, and faster results about employees' Internet usage behaviors.

**4.2 Log File Format and Data Analysis**

Firewall log file usually contains behavioral data about users when navigate Web sites to provide feedback about Internet usage for each employee in a company. This study is restricted to one type of log file format that is, Microsoft ISA 2004 log format because it is the most widely used log files format in business companies. The file record contains the following fields: User Address, Authenticated User, Client Agent, Authorization Status, Date, Time, Server Name, Proxy Name , Referring Server Name, Destination Name, : Destination IP, Destination Port, Processing Time, Bytes Sent, Bytes Received, Protocol Name, Transport , Operation , Object MIME, Source (Internet or cache) , and Result code.

The following is a fragment from a server logs file using ISA format:

172.29.10.1, James, -, Y, 2/4/96, 8:22:56, SERVERNAME, PROXYNAME, -, www.atw.fullfeed.com, -, 80, 5277, 4792, 890, http, TCP, GET, http://www.atw.fullfeed.com/, TEXT/HTML, Inet, 200,

- Field 1: User IP|Address = "172.29.10.1"
- Field 2: Authenticated User = " James "
- Field 3: Client Agent = " Mozilla/4.0 ….."
- Field 4: Authorization Status = " y"
- Field 5: Date ="2/4/96"
- Field 6: Time ="8:22:56"
- Field 7: Server Name ="ISA"
- Field 8: Proxy Name = " Websence "
- Field 9: Referring Server Name = " - "
- Field 10: Destination Name eg " www.atw.fullfeed.com "

- Field 11: Destination IP = = "10.29.10.1"
- Field 12: Destination Port ="80"
- Field 13: Processing Time= "5277"
- Field 14: Bytes Sent= "4792"
- Field 15: Bytes Received = "890"
- Field 16: Protocol Name = " http "
- Field 17: Transport =" TCP "
- Field 18: Operation =" GET "
- Field 19: Target Resource = "http://www.atw.fullfeed.com/,"
- Field 20: Object MIME =" TEXT/HTML "
- Field 21: Source (Internet or cache) " Inet "
- Field 22: Result Code =(0=No information available 5=Rejected 200=successful for Web proxy 304= successful for winsock proxy)

For the purpose of identifying Internet abuse, a sample Internet log file is analyzed 'with permission', and the results of analysis are presented statistically in a graphical format steps to analyze the data of the selected log file are presented next.

(1) A database structure is created and its fields are configured to records of MS ISA 2004 file content. MS ACESS 2003 is used for this purpose, and then the file records are mapped into matching fields in the database. The application which we have developed is designed in a way that it is able to read each record in the log file and transfer its content into the database format.

**(2)** The visited Websites (URLs) of the file are classified into categories. Each category contains all similar type of URLs in term of content or the service that Website provides. In this process, we first need to analyze the log file data to reduce the list of Web site addresses (i.e., URLs), because each hits in log file includes visited URL. For example www.google.com could appear in thousand numbers of hits, so we need to include only one entry for each unique site.

The URL visited by an employee is considered appropriate/inappropriate based on both the nature of business and job role of a given employee. For example, an employee working at the manufacturing department and has visited a financial Web site; this Website is considered inappropriate for this employee, but considered appropriate for an employee in the financial department and so forth (see table 1). Therefore, a special software (Microsoft ISA, 2004) is used to classify the URLs so that each URL belongs to appropriate or inappropriate category. The table below shows examples of appropriate and inappropriate URLs for a given organization.

Table (1): Sample URL categories

| Appropriates URL categories | Inappropriate URL categories |
|---|---|
| • Business and Economy<br>• Computer and Internet Info<br>• Educational Institutions<br>• government site<br>• Reference and Research | • Nudity<br>• Provocative Attire<br>• Politics/Opinion<br>• CS– Criminal Skills<br>• Entertainment<br>• Extreme |

**(3) Categorizing employees:** employees are categorized into groups according to common and shared attribute such as job duties, position, or job title and assign which tasks are appropriate /inappropriate for each group. For example, marketing groups include sale employee and marketing employee.

**(4) Determine criteria for Internet abuse (according to company policy):** this step is to determine the criteria for Internet abuse such as download size to determine consuming of bandwidth and network resources, type of downloaded file (ex. Video, audio, or exe file), and duration (time spent surfing the Internet). For example if an employee uses the Internet at the peak load (say between 9:00 and 11:00 am) and the Internet usage is not part of his job. This is considered as abuse because he/she uses the Internet at inappropriate time according to company policy. Also in this phase a criteria for each employee group is set based on their job and company policy. For example, maximum download size for, say, marketing group is 2 Mega Byte per month and the inappropriate time to log in into Internet between 10:00 AM to 11:30 AM, and so on.
In the process of determining the time that a user spend on the Internet (through using log file), a problem arises when, for example, a user start the session at 7:00 am, and only spent 15 minute and he forget to log off the Internet browser from 7:15 until 11:00 am, the user session appears in log file beginning at 7:00 and ends at 11:00 am. So when calculating the time which this user spent on the Internet, based on the difference between beginning time and end time it will be 3 hours, but actually he spends only 15 minutes. This problem was eliminated by calculating the beginning time and end time every 15 minutes to minimize probability of error in calculating time that user spent in browsing the Internet. If there is no hits in log file every 15 consecutive minutes we consider that the employee didn't use the Internet during this period.

**(5) Analysis of results:** The last step is to analyze the log file extracted data and generate statistical and graphical representation of results. The reports are produced in two levels. Level one reports are generated from the statistical analysis, and produces a graphical representation of the data after reorganizing and summarizing the data as mentioned earlier by using DBMS query commands. The output shown in the report represents URLs visited, size of data downloaded, time employee spent on the Internet for all employees in general, and for each individual employee. Level two reports will link this analysis with identified policy criteria of the organization in terms of specific time and date. Finally, the results in the two reports are compared to show if there is differences (discrepancy) in identifying Internet abuse.

### 4.3 A framework to manage employee Internet abuse

The proposed framework is enlighten the relationship between employee productivity with time spent, time wasting, efficacy of policy (both management policy and firewall policy) in addition to sites visited. The following decision table stub depicts main variables extracted from firewall log file and company policy, so that an organization management can take appropriate decision to prevent Internet usage abuse.

Table 1 Framework for making decision about Internet abuse

| **EMPLOYEE INFO.** | |
|---|---|
| Employee Id:<br>Job Title:<br>Department: | |
| **CONDITIONS** | **CONDITION VALUES** |
| Web site visited | Related to job description<br>Not related,<br>(i.e., values Y/N) |
| Type of data files downloaded | Related to job description<br>Not Related,<br>(values Y/N) |
| Time spent surfing the Internet | Higher than specified time<br>Not higher,<br>(values Y/N) |
| Time of day | Within allowed time<br>Not within allowed time,<br>(values Y/N) |
| Size of Download | Within given Quota<br>Outside given Quota,<br>(values Y/N) |
| **ACTION** | **Appropriate action**<br>**Inappropriate action** |

Application software will operate on the database to retrieve and analyze the data of each employee and make a decision about individual employee or group of employees to specify he/she is abusing or not abusing the internet.

## 5. Experimentation & Results

In this section, we have analyzed a private company true log file 'with permission'. Various analysis have been carried out on this file to identify the user behaviours on the Internet, it include overall company report for specific employees to show their behaviours. Using the developed application, the system administrators is able to view employee and company URLs entry day, time spent in Internet, download size, number of hits for each appropriate and inappropriate category. He/she can generate the same report for each employee with regard to the specified criteria, through the graphic representation. The results of the analysis are shown in figures 1 trough 9.

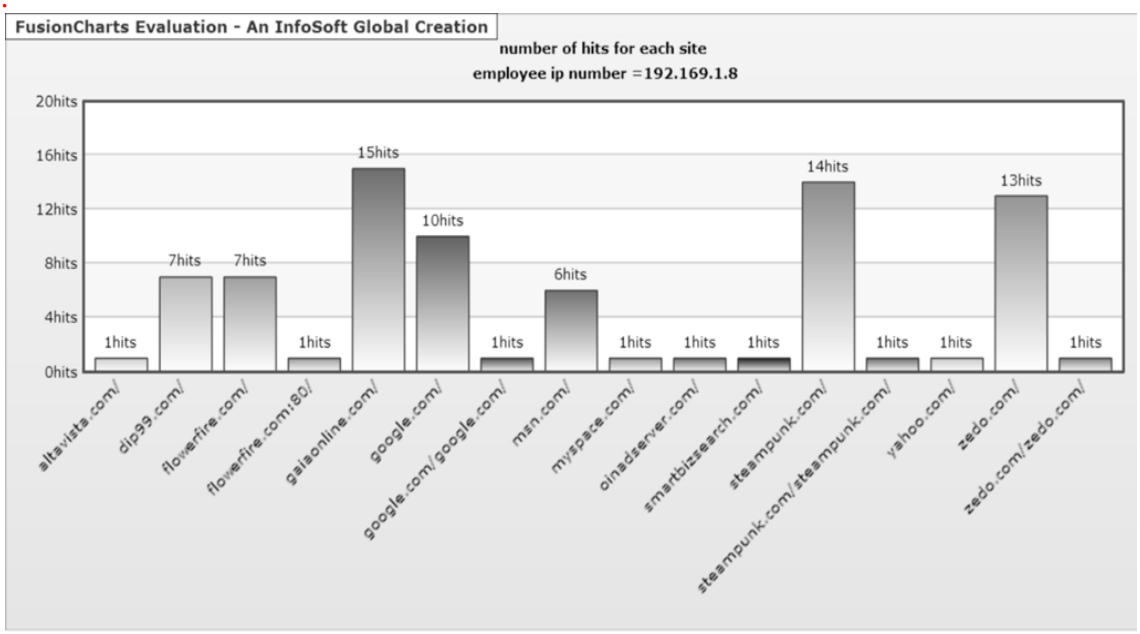


Figure (1): A sample URL analysis of an employee

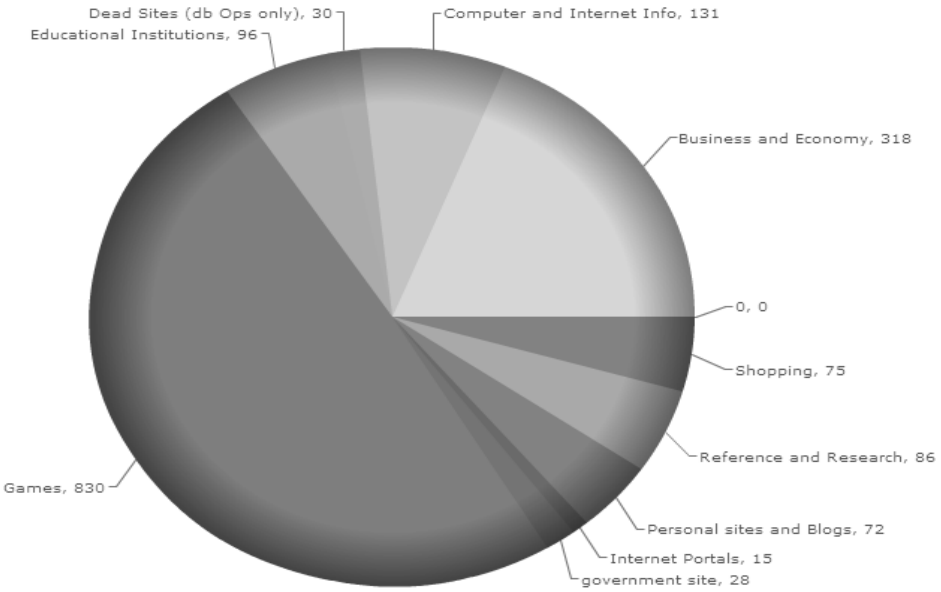


Figure (2): Number of hits for each entry URLs category for a specified period

Figure 1 represents number of hits for each entry URL in specific duration for a given employee. The graph in figure 2 shows the number of hits which performed by users (employees) according to URLs category. The graph in figure (3) represents downloading size for each user for related and unrelated work URLs categories.

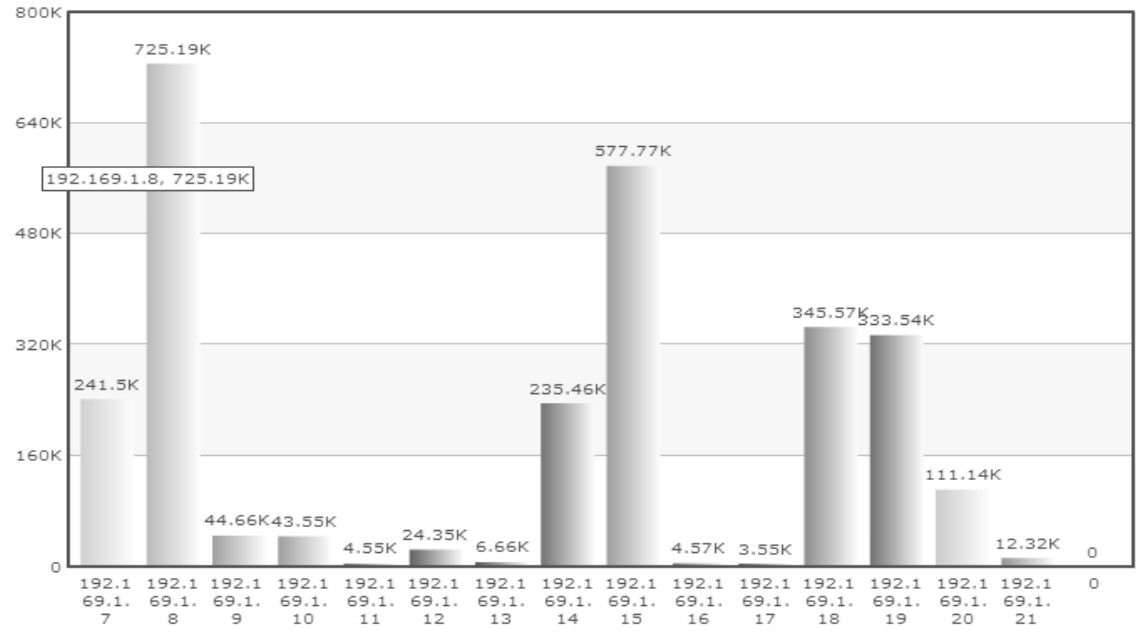


Figure (3): Sample of down loading size for each employee in the organization

The graph in figure (4) represents downloading size for each user employee based on only non-work related URLs category. This graph can also be developed for overall employees in an organization.

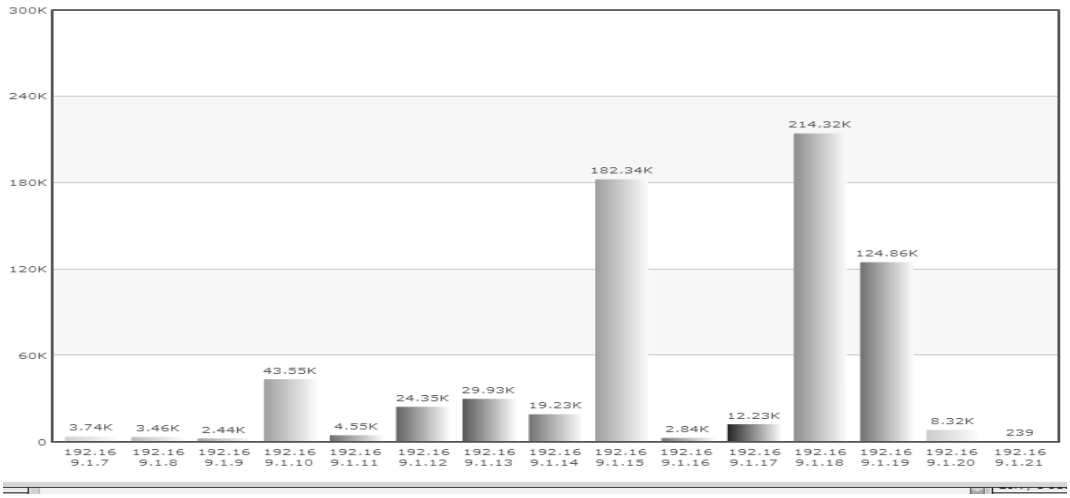


Figure (4): Download size for nonworking-related of the organizations' employees for a given period.

The graph in figure (5) represents the time that each user spent on using Internet within a specific period. Graph in Figure (6) represents the time that users spent on using Internet within specific date during workplace hours.

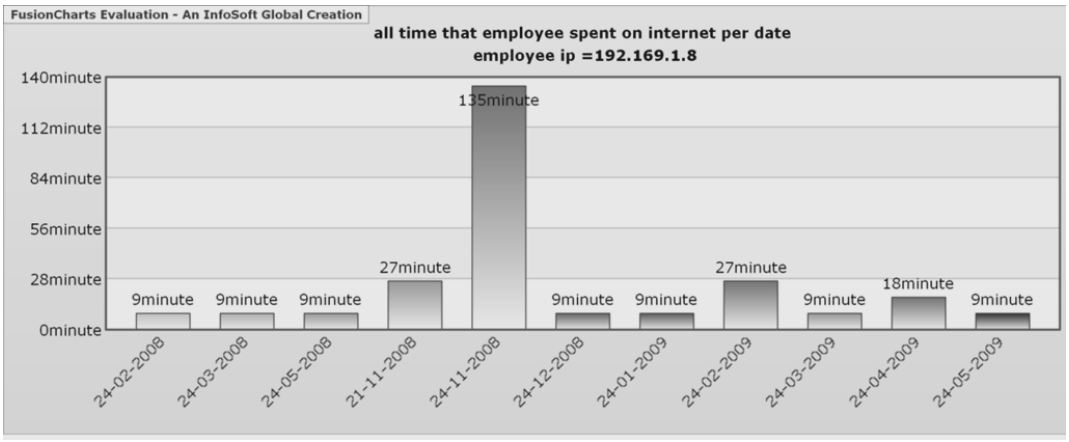


Figure (5): Time period users spent on the Internet within a specific period.

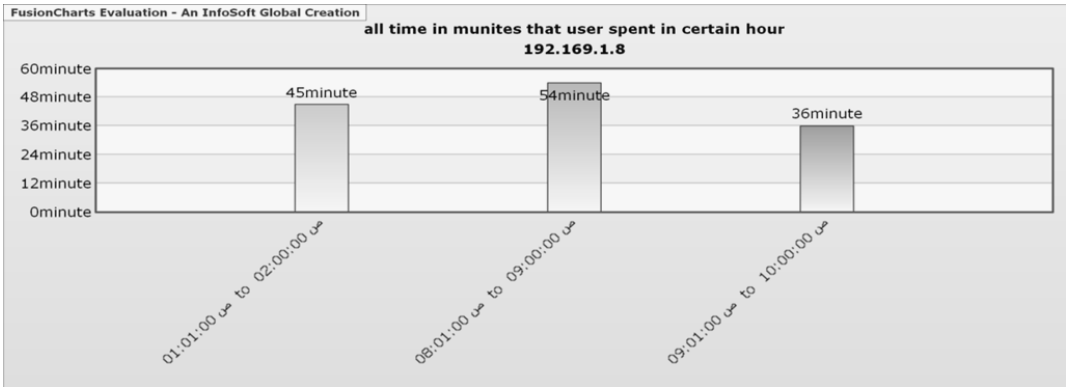


Figure (6): Time spent using the Internet for certain users

The graph in Figure (7) represents the download size for work related category for a specific employee in specific duration. The graph in figure (8) represents the number of hits for work related category for specific employee in specific duration.

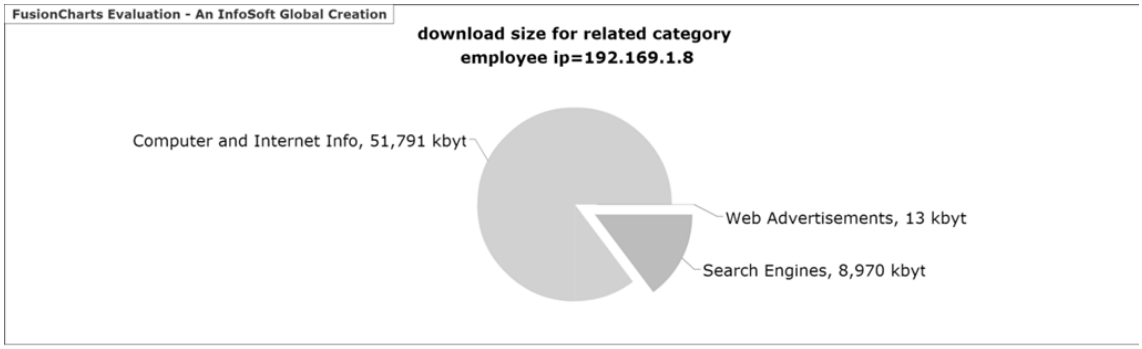


Figure (7): Work-related download size for an employee

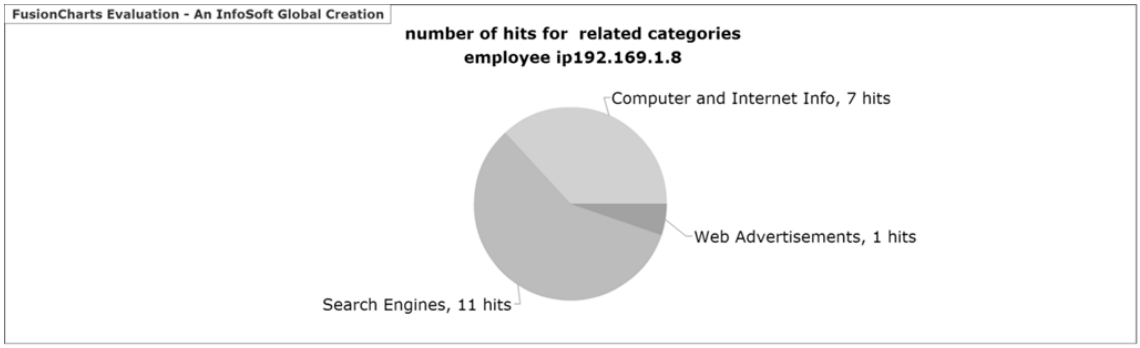


Figure (8): Number of hits for work related for specific employee in specific period

The graph in figure (9) represents the number of hits for nonworking related category for specific employee in specific duration.

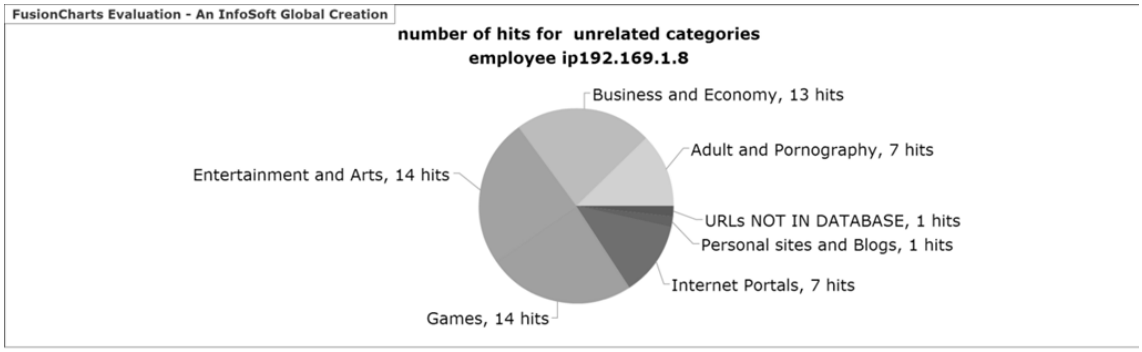


Figure (9)

## 6. Conclusions

Internet abuse and the associated consequences such as legal liabilities, negative publicity and excessive costs are considered main concerns across organizations worldwide. This research is to assist organizations in implementing effective corporate initiatives to improve employee Internet management practices. It is concerned with the analysis of firewall log files and associated organization policies to identify Internet abuse in an organization.

The results of analysing log files make identifying Internet usage behaviours more meaningful and more accurate especially when using external information behind log file raw data. Results could help in identifying employee performance and their productivity, and using this information by a company's management to develop Internet accepted use policy and to manage their employee use of the Internet to protect company resources, and probably for reward and punishment system. Through this analysis we can define the pattern of abuse in term of non-work related Web sites, extreme abuse, time spent on surfing the Internet, download size, and files type that user download it.

This technique of log file analysis is faced by many challenges including: large number of different log file format, size of log file (in a company of 500 users, one week's log file may contain more than one million records), log file data is limited and isolated from external information, diversity of log file data according to nature of business, implication raised by with the proliferation of mobile computing and wireless Internet appliances, and the ability to gain real log file for certain firms for reasons ranging from concern for employees' privacy to fear of revealing company proprietary information.

**References**

Akman, I., Mishra, A. (2010). Gender, age and income differences in Internet usage among employees in organizations. *Computers in Human Behavior* xxx, xxx–xxx,2010

American Management Association (2008). Electronic monitoring & surveillance survey: Over half of all employers combined fire workers for email and Internet abuse. American Management Association, March 13, 2008. Retrieved from http://press.amanet.org/press-releases/177/2007-electronic-monitoring surveillance-survey

Gee-Woo Bock and Swee Ling Ho (2009).Non-work related computing (NWRC). *Communications of the ACM*, vol. 52, no. 4, April (2009), pp. 124-128.

Carl. J. C., Young, K.S. (2004). Internet Abuse in the Workplace: New Trends in Risk Management, *Cyber psychology & Behavior*, Volume 7, Number 1

Christine, A., Kohut, G., Booth, R. (2009). Designing electronic use policies to enhance employee perceptions of fairness and to reduce cyberloafing: An empirical test of justice theory. *Computers in Human Behavior*, 25, pp. 902–910

Facca, F.M, Lanzi, P.L. (2005). Mining interesting knowledge from Weblogs: a survey. *Data & Knowledge Engineering*, Vol. 53, issue 2, pp. 225-241, June.

Flynn, N. (2005). *Electronic monitoring & surveillance survey*, American Management Association

Gaskin, J. (1998). Internet acceptable usage policies. *Information Systems Management*, 15, pp. 20–25

Gee-Woo Bock, Huei-Huan S., Kuan, G., Ping Liu, P., Sun, H. (2007). The Role of Task Characteristics and Organization Culture in Non-Work Related Computing (NWRC). *Human-Computer Interaction*, Part I, HCII 2007, LNCS 4550, pp. 681–690.

Gupta, J. (2004). Improving Workers' Productivity and Reducing Internet Abuse, *The Journal of Computer Information Systems*, 44, 2, pp. 74-78.

Johnson, J., Chalmers, K. W. (2007). Identifying Employee Internet Abuse, hicss, pp. 247b, *40th Annual Hawaii International Conference on System Sciences* (HICSS'07)

Lee, Z., Lee, Y., Kim, Y. (2004). Personal Web Page Usage in Organizations. In Anandarajan, M., Simmers C. A. (Eds.): Personal Web Usage in the Workplace: A guide to Effective Human Resources Management., *Information Sciences Publishing*, Hershey, PA, pp. 28-45.

Panko, R, Beh, H. (2002). Monitoring for Pornography and Sexual Harassment. *Communications of the ACM, 45.*

Sprnapshot Spy (2010). Employee Computer & Internet Abuse Statistics, A Virtual Imagination Inc publication Retrieved from http://www.snapshotspy.com/employee-computer-abuse-statistics.htm

Urbaczewski, A., Jessup L. M. (2002) Does Electronic Monitoring of Employee Internet Usage Work? *Communications of the ACM,* 45, 1, January.

Wen, J., chwieger, D. S., Gershuny, P (2007). Internet Usage Monitoring in the Workplace: Its Legal Challenges and Implementation Strategies. *Information Systems Management*, 24, pp. 185–196.

Young, K. (2010) Policies and procedures to manage employee Internet abuse. Computers in Human Behavior 26(6), pp. 1467-1471 (2010)

Young, K.S., Case, C. J. (2002). Employee Internet Management: Current Business Practices and Outcome. *CyberPsychology and Behavior*, 5(4), pp. 355-361

Young, K.S., Case, C. J. (2003) Employee internet abuse: Risk management strategies

and their effectiveness. *Proceedings of the American Society of Business and Behavioral Sciences*, Las Vegas, February 21, 2003, pp.1688-1694. Retrieved from http://www.netaddiction.com/articles/eia_strategies.pdf