Agricultural Research Center
Central Laboratory of Agricultural Expert Systems
RADCON Project, UTF/EGY/021

TR/RADCON/2007.4/Heggi.2

Contents

1. **Monitor and analyze the web usage of RADCON and VERCON servers including the most accessed components in the site, the measurement of web pages response time, and the recommendations for minimizing the accessing time if there is a need……..…..2**
   1.1 HTTP Construction Model…….…..3
   1.2 Traditional Web Characteristics…….…..3
   1.3 Web Applications Characteristics……4
   1.4 How RADCON Server can be measured through search engines……...…4
   1.5 Web Traffic Reporting Approaches……5
   1.6 Real Time Statistics Approach…….……...…13
   1.7 Network Monitor……..……..14
   1.8 Using Searching Engine to measure Web Traffic……..……15
   1.9 Conclusion

2. **Monitor and analyze the network security related events that occur on the network security elements such as Firewall and Intrusion Prevention System particularly that have an effect on RADCON servers and provide a methodology for monitoring process of valuable nodes in network, also provide an optimal configuration for the current situation……...18**
   2.1 1 Configuration Design……..……....18
   2.2 Analysis for the security events that occurs on RADCON/VERCON System……19
   2.3 IPS Log Analyses…….. 28
   2.4 Conclusion……29

3. **Design and monitoring the implementation of a data protection and recovery plan for RADCON Servers……..30**
   3.1 Specifications of RADCON Servers and backup storage…….……30
   3.2 Data protection plan……..30
   3.3 Implementation Plan…….……34

1. Monitor and analyze the web usage of RADCON and VERCON servers including the most accessed components in the site, the measurement of web pages response time, and the recommendations for minimizing the accessing time if there is a need.

This part is a comparative analysis among the different techniques to measure the quality of Web sites. There two major point of views to be mentioned when we try to evaluate a web site, performance and contents. In this part will investigate the different approaches that are used to measure both of them especially concerning the performance because of its impaction on web usage in spite of its valued content.

1.1 HTTP Construction Model

The actual traffic generated by HTTP is very complicated to be calculated, especially if the network path is included. To envision the operation of HTTP we construct a simple model which consists of a client and server. The client establishes a TCP connection to server, and then issues a request. The server processes the request and returns a response as indicated in figure 1.

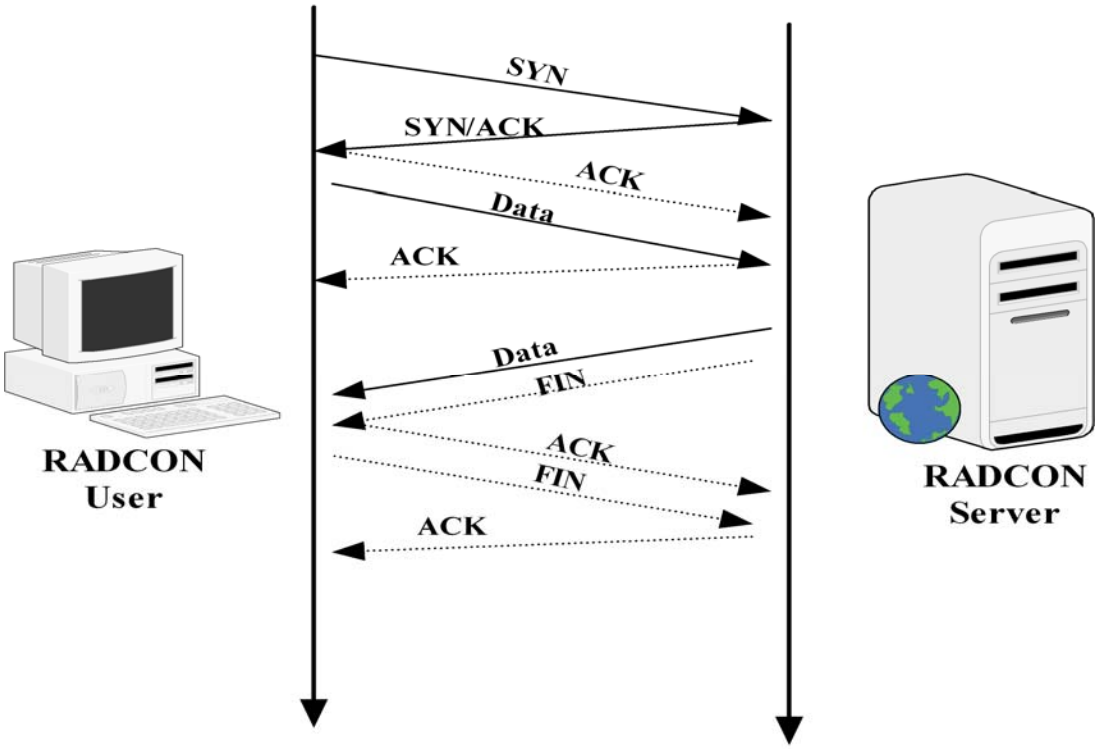

Figure 1: HTTP Operation

1.2 Traditional Web Characteristics

- **Hit rate:** page impression/view, visits, and visitors
- **User action:** entry site, exit path, entry click path, and exit click path
- **Techno graphical data:** operating system, browser, screen resolution, plug-ins, cookies, and pop-up-blocker
- **Top level domain and origin**

- **Stickiness:** how deep gets the user into page hierarchy, off-time between two visits

1.3 Web Applications Characteristics

Web application enables information processing to be done remotely from browser software and executed partly on a web server, application server and/or database server. Measuring the usability of web application is a very important researching area. Generally, there are four common factors that impact the usability of the web application and defined as the following:

1.3.1 Users

The users affected by the web interface can categorize the stakeholders of a web application. Users can be classified into primary users and secondary users. Primary user's class can be examined based on their competence, which change over time: novice, advanced beginners, and experts. If a category like novice user is important, then ease of use is an important usability attributes, whereas an expert user may require greater focus on efficient use. Users can perform transactions on a web application with varying levels of credentials. The level of authentication could be strong, moderate, weak, or unauthenticated.

1.3.2 Task

Task means the function of web application like informational, interactive, transaction, workflow, collaborative work environment, web portals, and web services. The type of the task and its complexity affects usability.

1.3.3 Technology

Development of web application can be for intranet, or Internet, different Internet transmission speeds, and device capabilities, therefore, the technological characteristics have a greater effect on web application.

1.3.4 Context

The contextual properties of a user that is interacting with a web application can vary with each web application. User context allows identification and enables personalization. Network provides network and bandwidth context. Location captures information about the location that can enhance context of web application. Industry context highlights special needs of an industry in relation to usability. Contextual properties, customization and industry classification provide the characteristics of web application that enable the environment to be tailored to the stakeholders.

1.4 How RADCON Server can be measured through search engines

Most search engines have three parts:

1. A crawler: wanders the web, following links and picking up information for its database. Crawlers do most of their work at times of the day when search engines are less busy, but they typically visit frequently updated pages more often.

2. **An index:** Once the crawler has collected all that text, it is then stored and indexed. This allows people searching for key words and phrases to get results relating to what they were searching for - their search results. Most sites will incorporate rating systems such as Google Page Rank or Alexa Rankings in positioning your site. These ratings are used attempt to ensure that sites that are important receive more traffic than unimportant sites.

3. **Interface:** Search engines provide a public interface for users who want to find information on the web. They can type the word or phrase they're searching for and the interface will run an algorithm to find the pages relevant to their search and display them. The most popular search engines today include Google, Yahoo, and MSN.

1.5 Web Traffic Reporting Approaches

1.5.1 Log File Analyzer Approach

A web server creates a record of the traffic and information requests in log files. These log files include information on errors, processing time, bandwidth used, visitor IP address, where visitors came from (referred) along with additional information such as operating system or browser used.

1.5.2 Examples of log file reporting applications

- **Web trends:** www.webtrends.com **(commercial)**
- **Saw mill:** www.sawmill.net **(commercial)**
- **Analog:** www.analog.cx **(free)**
- **Webalizer:** www.webalizer.org **(free)**

1.5.3 Notes on VERCON web site log files as an example

Web server logs are plain text (ASCII) files, independent of server platform. There are some differences between server software, but traditionally there are four types of server logs:

- Transfer (access) log
- Error log
- Referrer log
- Agent log

The first two types of log files are standard. The referrer and agent logs may or may not be “turned on” at the server or may be added to the transfer log file to create an “extended” log file format. Each HTTP protocol transaction, whether completed or not, is recorded in the logs.

Transfer (access) log

The following is an example of a single line in a common transfer log. This typically displays as one long line of ASCII text, separated by tabs and spaces.

2007-04-16 00:00:18 W3SVC22260 FRONTEND 192.168.1.16 GET /extpub/main.asp PageNo=2&vuserid=7870 80 - 62.135.62.18 HTTP/1.1 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1) ;+VerconUserID=7870;+Snitz00User=Cookies=&Pword=0101655192&Name=samywageh - www.vercon.sci.eg 200 0 121 0 526 399531

2007-04-16: date
00:00:18*:* time
W3SVC22260*:* service name
FRONTEND: server name
192.168.1.16: server ip address
GET: Method
/extpub/main.asp*:* the resource accessed
PageNo=2&vuserid=7870*:* user query
80: server port
62.135.62.18: This is the address of the computer making the HTTP request.
HTTP/1.1: protocol version
Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1): the browser used by the client, it is called a user agent
+VerconUserID=7870;+Snitz00User=Cookies=&Pword=0101655192&Name=samywageh: cookies
www.vercon.sci.eg: The site that directed the user to the current site
200: The request was fulfilled without error
121: The number of bytes that the server sent to the client
526: The number of bytes that the client sent to the server
399531: the length of time the action took to complete. Time Taken is logged in milliseconds, according to the following technical breakdown: The client-request timestamp is initialized when HTTP.sys (the kernel-mode driver) receives the first byte (before HTTP.sys begins parsing the request). The client-request timestamp is stopped when the send completion occurs (for the last send) in IIS. Time Taken does not reflect time across the network. Also note, the first request to the site shows a slightly longer time taken than other similar/same requests because HTTP.sys opens the log file with the first request.

Hits, Views, and Visits

Hit counters continue to be popular features on web pages, but they, in fact, have little value. First, most hit counters can be adjusted to start at any number. So any number you see in on a hit counter may be artificial. Second, just what is defined as a hit? In fact, requesting a single web page can result in multiple hits to the server. The page of html will show as a hit and each graphic on the page will also record as a hit in the log. So a page of html with six graphics in a navigation bar could record eight individual hits in the log. This set of eight hits is described as a view, all the hits necessary to display a web page. The next analysis is to look through the views to recreate the user's visit to the site – how they got to your site, where they went in the site and how long they spent.

1.5.4 Example for using of Usage Statistics for vercon.sci.eg by Log Analysis Approach

The statistics especially concerning the period from Oct., 2006 to end or Mar., 2007 is described in table 1. These statistics indicate the average during 6 months is 4127456 Kbytes which equivalent to 4 GB traffic per month on average basis.

Table 1: Summary of VERCON Statistics in the period from 1/10/2006 to 31/3/2007

Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Mar 2007	7698	4702	2087	352	6853	4718776	10939	64720	145767	238652
Feb 2007	7015	4376	2004	296	5186	3783057	8302	56130	122546	196432
Jan 2007	6396	4302	1924	276	5208	3656277	8568	59673	133365	198278
Dec 2006	9929	6784	2950	297	5566	4515150	9231	91469	210315	307800
Nov 2006	10299	6997	3088	322	5932	4936308	9679	92651	209919	308972
Oct 2006	5273	3607	1481	228	4338	3155169	7081	45911	111847	163483
Totals						24764737	53800	410554	933759	1413617

Table 2 indicates the statistics concerned with March 2007 in details. The most important comments on this table are the following:

- Maximum visit per day is 436, if we consider a visit is a user so, 436 is the maximum number of users visit the site per day
- Maximum traffic per day is 0.2 GB
- The number of response code that give a file not found is 15318 from successful response 145767 which represents 10.5 percent, therefore, these not found links in the site should be reviewed

Table 2: VERCON Web Site Statistics on March 2007

Monthly Statistics for March 2007		
Total Hits		**238652**
Total Files		**145767**
Total Pages		**64720**
Total Visits		**10939**
Total Kbytes		**4718776**
Total Unique Sites		**6853**
Total Unique URLs		**3230**
Total Unique Referrers		**894**
Total Unique User Agents		**927**
	Avg	**Max**
Hits per Hour	**320**	**3503**
Hits per Day	**7698**	**12828**
Files per Day	**4702**	**7008**
Pages per Day	**2087**	**3738**
Visits per Day	**352**	**436**
KBytes per Day	**152219**	**216205**
Hits by Response Code		
Undefined response code		**545**
Code 200 - OK		**145767**
Code 206 - Partial Content		**1354**
Code 301 - Moved Permanently		**17**
Code 302 - Found		**6027**
Code 304 - Not Modified		**67404**
Code 400 - Bad Request		**52**
Code 401 - Unauthorized		**5**
Code 403 - Forbidden		**264**
Code 404 - Not Found		**15318**
Code 406 - Not Acceptable		**1**
Code 500 - Internal Server Error		**1898**

Table 3 indicates the daily statistics for VERCON web site. The most important comments on this table are:

- The maximum number of visits is 463
- The minimum number of visits is 272
- The maximum amount of traffic per day is 0.2 GB

Table3: Daily VERCON Site Statistics

Daily Statistics for March 2007												
Day	**Hits**		**Files**		**Pages**		**Visits**		**Sites**		**KBytes**	
1	**8495**	3.56%	**5406**	3.71%	**2324**	3.59%	**393**	3.59%	**343**	5.01%	**144836**	3.07%
2	**6006**	2.52%	**4086**	2.80%	**1432**	2.21%	**340**	3.11%	**276**	4.03%	**135470**	2.87%
3	**6790**	2.85%	**4248**	2.91%	**1656**	2.56%	**353**	3.23%	**320**	4.67%	**166014**	3.52%
4	**11598**	4.86%	**6658**	4.57%	**3131**	4.84%	**436**	3.99%	**353**	5.15%	**216205**	4.58%
5	**7414**	3.11%	**4433**	3.04%	**2101**	3.25%	**374**	3.42%	**337**	4.92%	**175808**	3.73%
6	**8322**	3.49%	**4949**	3.40%	**2272**	3.51%	**392**	3.58%	**336**	4.90%	**177889**	3.77%
7	**7905**	3.31%	**4475**	3.07%	**2069**	3.20%	**356**	3.25%	**303**	4.42%	**157667**	3.34%
8	**7577**	3.17%	**4106**	2.82%	**2104**	3.25%	**347**	3.17%	**314**	4.58%	**110782**	2.35%
9	**4601**	1.93%	**3017**	2.07%	**1212**	1.87%	**272**	2.49%	**256**	3.74%	**119398**	2.53%
10	**5381**	2.25%	**3463**	2.38%	**1493**	2.31%	**277**	2.53%	**257**	3.75%	**117142**	2.48%
11	**12828**	5.38%	**7008**	4.81%	**3738**	5.78%	**353**	3.23%	**321**	4.68%	**165230**	3.50%
12	**11720**	4.91%	**5730**	3.93%	**3333**	5.15%	**344**	3.14%	**341**	4.98%	**159794**	3.39%
13	**7819**	3.28%	**4656**	3.19%	**2260**	3.49%	**346**	3.16%	**319**	4.65%	**123496**	2.62%
14	**7815**	3.27%	**4858**	3.33%	**2076**	3.21%	**374**	3.42%	**363**	5.30%	**147802**	3.13%
15	**7234**	3.03%	**4121**	2.83%	**2054**	3.17%	**336**	3.07%	**293**	4.28%	**151057**	3.20%
16	**6399**	2.68%	**3531**	2.42%	**2217**	3.43%	**309**	2.82%	**280**	4.09%	**121882**	2.58%
17	**6177**	2.59%	**4321**	2.96%	**1492**	2.31%	**375**	3.43%	**331**	4.83%	**160082**	3.39%
18	**7404**	3.10%	**4787**	3.28%	**2053**	3.17%	**394**	3.60%	**363**	5.30%	**172131**	3.65%
19	**9129**	3.83%	**6087**	4.18%	**2113**	3.26%	**392**	3.58%	**346**	5.05%	**182296**	3.86%
20	**9232**	3.87%	**5735**	3.93%	**2657**	4.11%	**390**	3.57%	**355**	5.18%	**163702**	3.47%
21	**7498**	3.14%	**4724**	3.24%	**1802**	2.78%	**354**	3.24%	**313**	4.57%	**159969**	3.39%
22	**8096**	3.39%	**5189**	3.56%	**2257**	3.49%	**352**	3.22%	**312**	4.55%	**171878**	3.64%
23	**5108**	2.14%	**3334**	2.29%	**1513**	2.34%	**258**	2.36%	**237**	3.46%	**116286**	2.46%
24	**7933**	3.32%	**4907**	3.37%	**1935**	2.99%	**324**	2.96%	**308**	4.49%	**149666**	3.17%
25	**7497**	3.14%	**4711**	3.23%	**1676**	2.59%	**341**	3.12%	**308**	4.49%	**141921**	3.01%
26	**7158**	3.00%	**4661**	3.20%	**1775**	2.74%	**327**	2.99%	**296**	4.32%	**161022**	3.41%
27	**7547**	3.16%	**4574**	3.14%	**2252**	3.48%	**345**	3.15%	**304**	4.44%	**166747**	3.53%
28	**7405**	3.10%	**4617**	3.17%	**2054**	3.17%	**388**	3.55%	**351**	5.12%	**143200**	3.03%
29	**9370**	3.93%	**5980**	4.10%	**2721**	4.20%	**376**	3.44%	**334**	4.87%	**150843**	3.20%
30	**4489**	1.88%	**2997**	2.06%	**1226**	1.89%	**385**	3.52%	**308**	4.49%	**129207**	2.74%
31	**6705**	2.81%	**4398**	3.02%	**1722**	2.66%	**355**	3.25%	**318**	4.64%	**159352**	3.38

Table 4 indicates the hourly statistics for VERCON web site. The most important comments on this table is that the high load period is between 8 a.m. and 4 p.m. which is the working hours of ARC users.

Table 4: Hourly VERCON web site statistics

Hourly Statistics for March 2007												
Hour	Hits			Files			Pages			KBytes		
	Avg	Total		Avg	Total		Avg	Total		Avg	Total	
0	100	3125	1.31%	69	2155	1.48%	28	876	1.35%	3102	96151	2.04%
1	72	2253	0.94%	52	1625	1.11%	16	519	0.80%	2845	88209	1.87%
2	34	1057	0.44%	23	733	0.50%	8	265	0.41%	997	30901	0.65%
3	45	1415	0.59%	28	878	0.60%	14	447	0.69%	1841	57072	1.21%
4	69	2152	0.90%	45	1423	0.98%	19	603	0.93%	1776	55041	1.17%
5	100	3121	1.31%	63	1962	1.35%	28	879	1.36%	2052	63603	1.35%
6	200	6210	2.60%	110	3420	2.35%	60	1877	2.90%	3285	101840	2.16%
7	500	15516	6.50%	269	8354	5.73%	152	4729	7.31%	5427	168243	3.57%
8	816	25312	10.61%	421	13064	8.96%	236	7319	11.31%	7859	243633	5.16%
9	834	25855	10.83%	472	14648	10.05%	242	7530	11.63%	11346	351733	7.45%
10	636	19721	8.26%	353	10963	7.52%	177	5508	8.51%	8188	253832	5.38%
11	604	18736	7.85%	347	10773	7.39%	184	5714	8.83%	9613	298010	6.32%
12	351	10887	4.56%	225	6995	4.80%	92	2863	4.42%	8025	248787	5.27%
13	294	9139	3.83%	201	6234	4.28%	74	2296	3.55%	7474	231706	4.91%
14	290	9011	3.78%	201	6247	4.29%	76	2372	3.67%	7054	218682	4.63%
15	243	7540	3.16%	164	5094	3.49%	61	1897	2.93%	5968	185012	3.92%
16	304	9440	3.96%	209	6497	4.46%	69	2150	3.32%	7355	227990	4.83%
17	340	10547	4.42%	234	7259	4.98%	78	2434	3.76%	9045	280382	5.94%
18	393	12193	5.11%	258	8020	5.50%	87	2713	4.19%	10332	320292	6.79%
19	378	11727	4.91%	256	7949	5.45%	78	2439	3.77%	10105	313256	6.64%
20	399	12376	5.19%	244	7583	5.20%	122	3790	5.86%	10410	322696	6.84%
21	277	8599	3.60%	187	5810	3.99%	71	2204	3.41%	7174	222409	4.71%
22	223	6924	2.90%	139	4329	2.97%	55	1715	2.65%	5817	180317	3.82%
23	186	5796	2.43%	121	3752	2.57%	51	1581	2.44%	5128	158978	3.37%

Table 5 indicates the top ten accessed parts of VERCON site, this table indicates that the most used part of the system is the part concerning the extension centers then extension publications then forum. But this statistics needs more elaboration to identify to which subsystem the part is belonged. Moreover, the entry and exit pages are indicator to the important parts for users that lead them to access the site.

Table 5: top 10 accessed parts in site, entry pages, and exit pages

Top 10 of 3230 Total URLs By KBytes					
#	Hits		Kbytes		URL
1	607	0.25%	458582	9.72%	/ershad_centeral/Activites2002.htm
2	549	0.23%	343349	7.28%	/ershad_centeral/Activites2001.htm
3	491	0.21%	327161	6.93%	/ershad_centeral/active.htm
4	533	0.22%	307205	6.51%	/indexUI/uploaded/Mangoproduction/mangoproduction.htm
5	1418	0.59%	146137	3.10%	/extpub/main.asp
6	9024	3.78%	140530	2.98%	/vercon.asp
7	360	0.15%	85686	1.82%	/ershad_centeral/centers.htm
8	329	0.14%	66553	1.41%	/forumn/search.asp
9	71	0.03%	60670	1.29%	/indexUI/uploaded/Potatoproduction2005940/potaproduction2005.htm
10	95	0.04%	50571	1.07%	/indexUI/uploaded/Aranep+production/aranep+production.htm

Top 10 of 305 Total Entry Pages					
#	Hits		Visits		URL
1	9024	3.78%	3801	36.80%	/vercon.asp
2	607	0.25%	389	3.77%	/ershad_centeral/Activites2002.htm
3	533	0.22%	376	3.64%	/indexUI/uploaded/Mangoproduction/mangoproduction.htm
4	360	0.15%	304	2.94%	/ershad_centeral/centers.htm
5	491	0.21%	281	2.72%	/ershad_centeral/active.htm
6	549	0.23%	262	2.54%	/ershad_centeral/Activites2001.htm
7	1418	0.59%	251	2.43%	/extpub/main.asp
8	273	0.11%	227	2.20%	/aerdri/corpora.htm
9	458	0.19%	208	2.01%	/text12.asp
10	239	0.10%	178	1.72%	/aerdri/compinsi.htm

Top 10 of 518 Total Exit Pages					
#	Hits		Visits		URL
1	9024	3.78%	2444	23.68%	/vercon.asp
2	607	0.25%	390	3.78%	/ershad_centeral/Activites2002.htm
3	533	0.22%	377	3.65%	/indexUI/uploaded/Mangoproduction/mangoproduction.htm
4	549	0.23%	321	3.11%	/ershad_centeral/Activites2001.htm
5	491	0.21%	299	2.90%	/ershad_centeral/active.htm
6	360	0.15%	267	2.59%	/ershad_centeral/centers.htm
7	458	0.19%	260	2.52%	/text12.asp
8	1418	0.59%	221	2.14%	/extpub/main.asp
9	273	0.11%	204	1.98%	/aerdri/corpora.htm
10	1020	0.43%	195	1.89%	/forumn/topic.asp

Table 6 indicates the sites that refer to user to log in VERCON site. We note that most of the referrer sites is the main site itself or a subsystem included in the site.

Table 6: Top 30 Referrers

#	Hits		Referrer
1	37923	15.89%	- (Direct Request)
2	27130	11.37%	http://www.vercon.sci.eg/
3	17435	7.31%	http://www.vercon.sci.eg/vercon.asp
4	17011	7.13%	http://www.vercon.sci.eg/ershad_centeral/centers/extpages.asp
5	5218	2.19%	http://www.dvd4arab.com/showthread.php
6	4853	2.03%	http://www.vercon.sci.eg/insys/rsm/main.asp
7	4670	1.96%	http://www.vercon.sci.eg/extpub/main.asp
8	4069	1.70%	http://www.vercon.sci.eg/forumn/topic.asp
9	3307	1.39%	http://www.vercon.sci.eg/forumn/default1.asp
10	3085	1.29%	http://www.vercon.sci.eg/VerconProject/ExDoc5.asp
11	2801	1.17%	http://www.vercon.sci.eg/INSYS/Moderates/Asyaot/Asyaot.asp
12	2683	1.12%	http://www.vercon.sci.eg/forumn/register.asp
13	2530	1.06%	http://www.vercon.sci.eg/farm_prob/ProbSearch.asp
14	2514	1.05%	http://www.vercon.sci.eg/forumn/forum.asp
15	2484	1.04%	http://www.google.com.eg/search
16	2451	1.03%	http://dvd4arab.com/showthread.php
17	2410	1.01%	http://www.vercon.sci.eg/ershad_centeral/Activites2001.htm
18	2402	1.01%	http://www.vercon.sci.eg/ershad_centeral/Activites2002.htm
19	1919	0.80%	http://www.vercon.sci.eg/insys/rsm/researchstat.asp
20	1738	0.73%	http://www.vercon.sci.eg/farm_entrynew/main.asp
21	1686	0.71%	http://www.vercon.sci.eg/INSYS/Moderates/MainMod.asp
22	1678	0.70%	http://www.vercon.sci.eg
23	1660	0.70%	http://www.vercon.sci.eg/extpub/mainc.asp
24	1603	0.67%	http://www.vercon.sci.eg/ershad_centeral/View_Centers.asp
25	1479	0.62%	http://www.vercon.sci.eg/aerdri/corpora.htm
26	1458	0.61%	http://www.vercon.sci.eg/tomato/tomatomain.asp
27	1433	0.60%	http://www.vercon.sci.eg/ProbEntryPubNew/main_sec.asp
28	1314	0.55%	http://www.vercon.sci.eg/wheat2005/es.aspx
29	1275	0.53%	http://www.vercon.sci.eg/Vercon_en/vercon.asp
30	1221	0.51%	http://www.vercon.sci.eg/aerdri/progersh.htm

Table 7: Top 30 Countries

#	Hits		Files		KBytes		Country
1	152188	63.77%	92108	63.19%	2915473	61.78%	Unresolved
2	58702	24.60%	34166	23.44%	1094281	23.19%	Network
3	13060	5.47%	8327	5.71%	289105	6.13%	Egypt
4	7055	2.96%	4616	3.17%	133549	2.83%	US Commercial
5	1857	0.78%	1730	1.19%	71509	1.52%	United Arab Emirates
6	1723	0.72%	1520	1.04%	71951	1.52%	Saudi Arabia
7	1207	0.51%	1072	0.74%	50242	1.06%	Morocco
8	993	0.42%	657	0.45%	38404	0.81%	Syria
9	214	0.09%	178	0.12%	9895	0.21%	Germany
10	165	0.07%	147	0.10%	3838	0.08%	Austria
11	158	0.07%	134	0.09%	4300	0.09%	Israel

Top 30 Countries (Continued)							
12	**154**	0.06%	**85**	0.06%	**3013**	0.06%	United Kingdom
13	**136**	0.06%	**136**	0.09%	**1666**	0.04%	Japan
14	**129**	0.05%	**129**	0.09%	**8578**	0.18%	Oman
15	**116**	0.05%	**112**	0.08%	**1937**	0.04%	Jordan
16	**73**	0.03%	**71**	0.05%	**1842**	0.04%	Czech Republic
17	**69**	0.03%	**44**	0.03%	**725**	0.02%	France
18	**68**	0.03%	**57**	0.04%	**2281**	0.05%	Lebanon
19	**51**	0.02%	**48**	0.03%	**834**	0.02%	Netherlands
20	**50**	0.02%	**36**	0.02%	**240**	0.01%	Norway
21	**47**	0.02%	**44**	0.03%	**1412**	0.03%	Italy
22	**46**	0.02%	**22**	0.02%	**283**	0.01%	Finland
23	**43**	0.02%	**36**	0.02%	**2273**	0.05%	Old style Arpanet (arpa)
24	**41**	0.02%	**27**	0.02%	**1107**	0.02%	US Educational
25	**38**	0.02%	**33**	0.02%	**888**	0.02%	Canada
26	**36**	0.02%	**26**	0.02%	**2687**	0.06%	Non-Profit Organization
27	**27**	0.01%	**25**	0.02%	**966**	0.02%	Romania
28	**22**	0.01%	**21**	0.01%	**940**	0.02%	Greece
29	**21**	0.01%	**16**	0.01%	**887**	0.02%	Malaysia
30	**17**	0.01%	**15**	0.01%	**88**	0.00%	Dominican Republic

Table 7 indicated the 30 top countries; it is very obvious that Egypt seizes the most percentage of accessibility. Also, there is a comment here in the first and second record, unresolved means that the IP of the host that enter the site doesn't has a DNS record, and the same situation has been occurred for the second record which is network I propose that the reason of this behavior is due to two reasons the IP address of ARC client that does not has a DNS entry and the request between Web server and database server and they represent unresolved and network entries respectively.

1.6 Real Time Statistics Approach

Another method of analyzing your Web site activity is to update an optimized reporting database each time a visitor comes to your site. This method requires you to place a small JavaScript on every web page that is to be tracked. This code is invisible to your site visitors. As visitors surf your site, the code places a cookie on their computer so they have a unique identifier and can be tracked. Within seconds of placing the JavaScript on the Web pages to be tracked, site visitor information is securely written out to a database and is instantly available for reporting. Once the data is captured, the optimized database engine aggregates and manipulates the data to report on areas such as what marketing campaign referred the visitor to your site or such data as order values. Because the data is real time, you do not have to wait for a snapshot report.

My comment in this method that it is similar to the first one except for the way it generates the report.

1.7 Network Monitor: packet sniffing is installed on each web server

In this method we can measure the amount of traffic used in server in the level of used protocol such as HTTP, HTTPS, SMTP, and others. We can utilize this measure in defining the actual consumed resources on server.

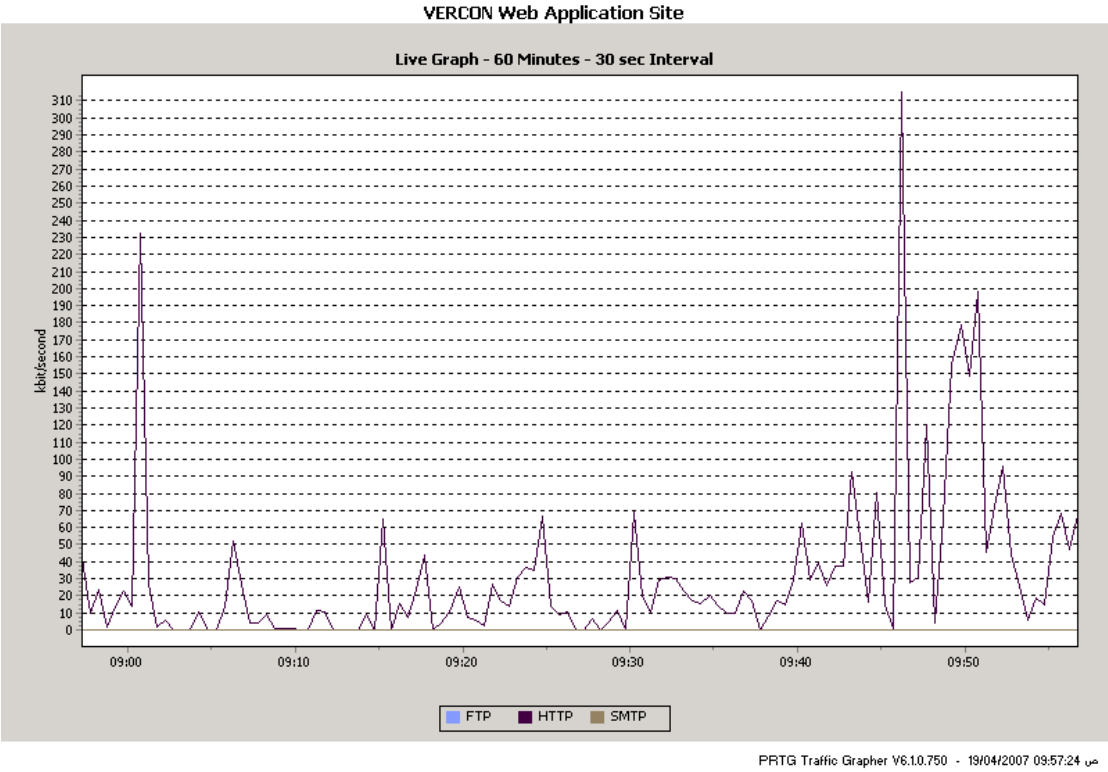

Figure 2.a: Monitor traffic for VERCON Web Server using Packet Sniffing during one hour

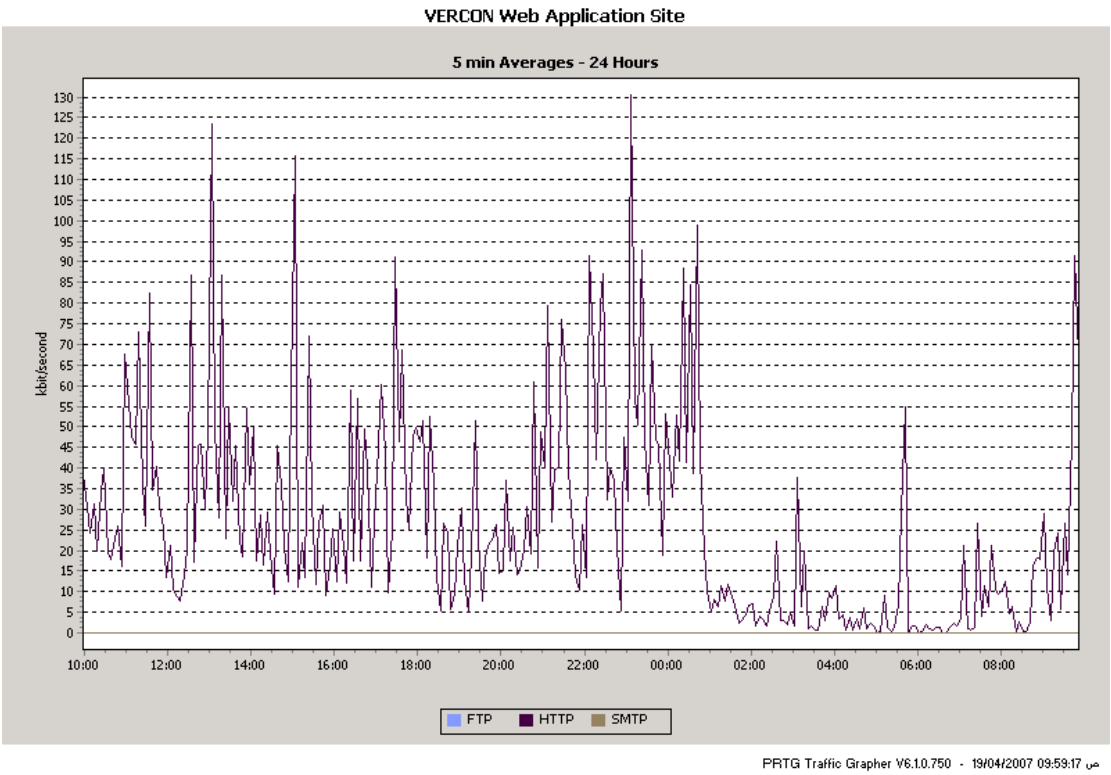

Figure 2.b: Monitor traffic for VERCON Web Server using Packet Sniffing during 1 Day

Figure 2: Monitor traffic for VERCON Web Server using Packet Sniffing

This method is suitable for identifying the consumed traffic at the level of different protocols and the data rate values in different times, therefore, it could be useful for estimation of used network resources by this site.

1.8 Using Searching Engine to measure Web Traffic

We use Alexa tools as an example to measure the traffic if VERCON site. Alexa computes traffic rankings by analyzing the Web usage of millions of Alexa Toolbar The traffic rank is based on three months of aggregated historical traffic data from millions of Alexa Toolbar users and is a combined measure of page views and users (reach). As a first step, Alexa computes the reach (measures the number of users as indicated in figure 3) and number of page views (measure the number of pages viewed by Alexa Toolbar users. Multiple page views of the same page made by the same user on the same day are counted only once as indicated in figure 4) for all sites on the Web on a daily basis. The main Alexa traffic rank is based on the geometric mean of these two quantities averaged over time (so that the rank of a site reflects both the number of users who visit that site as well as the number of pages on the site viewed by those users). The information is sorted, counted, and computed.

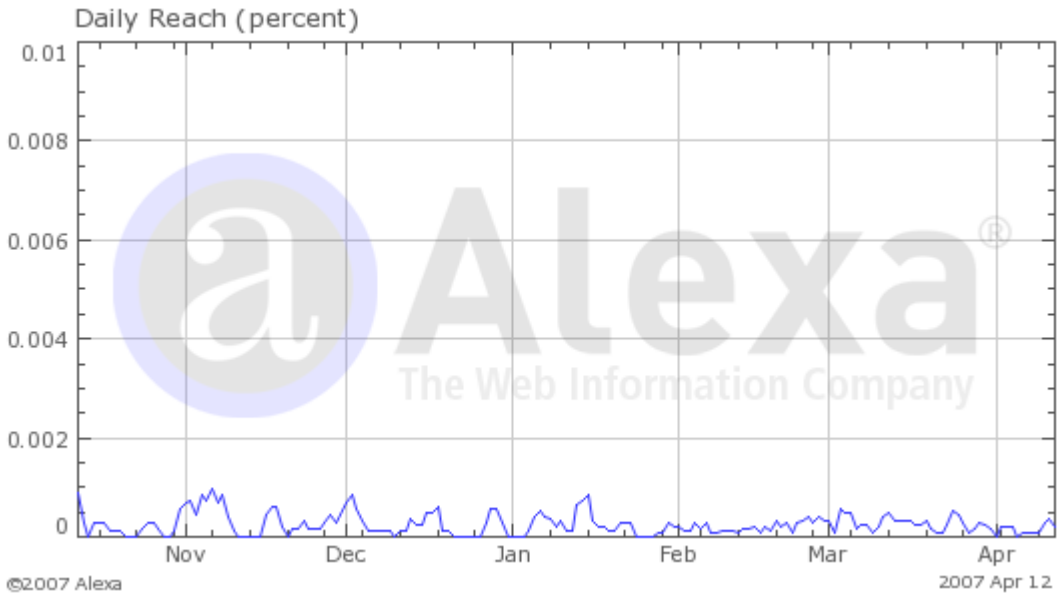

Figure 3: Daily Reach for VERCON Site using Alexa Search Engine

Figure 3 indicate that the average percent for the number of users that access VERCON site during three months is 0.00023% from Internet users. The number of unique pages viewed per user per day for this site is 3.3. Alexa traffic rank based on a combined measure of page views and users (reach) is 387195 during three months
Vercon.sci.eg users come from these countries:
Egypt : 42.4%
Saudi Arabia: 12.1%
Jordan: 9.1%
Kuwait: 9.1%
United Arab Emirates: 9.1%

1.9 Conclusion

From the previous mentioned different approaches to measure the metrics of Web applications we conclude the following:

- The selection for metrics of web application site is based on site objectives for example web business applications, response time metric is very important and so.
- Two parameters related to performance are not considered in the previous mentioned approaches: web server processing capabilities and security attacks that occurs on server but they can be considered to be ignored because the high capabilities of servers and the installed security system
- The selection of the approach to be used in measuring is related to which metrics are selected to be measured.
- For RADCON/VERCON web systems, user behaviors during access the system is not measured in the previous mention approaches so; other approaches should be adapted or used to measure the very vital metric.

- The HTTP protocol is stateless: web server only sees requests from some remote IP address. The remote address connects, sends a request, receives a response and then disconnects. The web server has no idea what the remote side is doing between these requests, or even what it did with the response sent to it. This makes it impossible to determine things like how long a user spends on your site.
- It should now be obvious that there are only certain things that can determine from a web server log such as how many requests generated a 404 (not found) result. Also, there are some wildly inaccurate and misleading numbers you can collect depending on what assumptions you make.
- The most accurate way to get an accurate picture of what web server is doing is to look at its logs.

2. Monitor and analyze the network security related events that occur on the network security elements such as Firewall and Intrusion Prevention System particularly that have an effect on RADCON servers and provide a methodology for monitoring process of valuable nodes in network, also provide an optimal configuration for the current situation.

2.1 Configuration Design

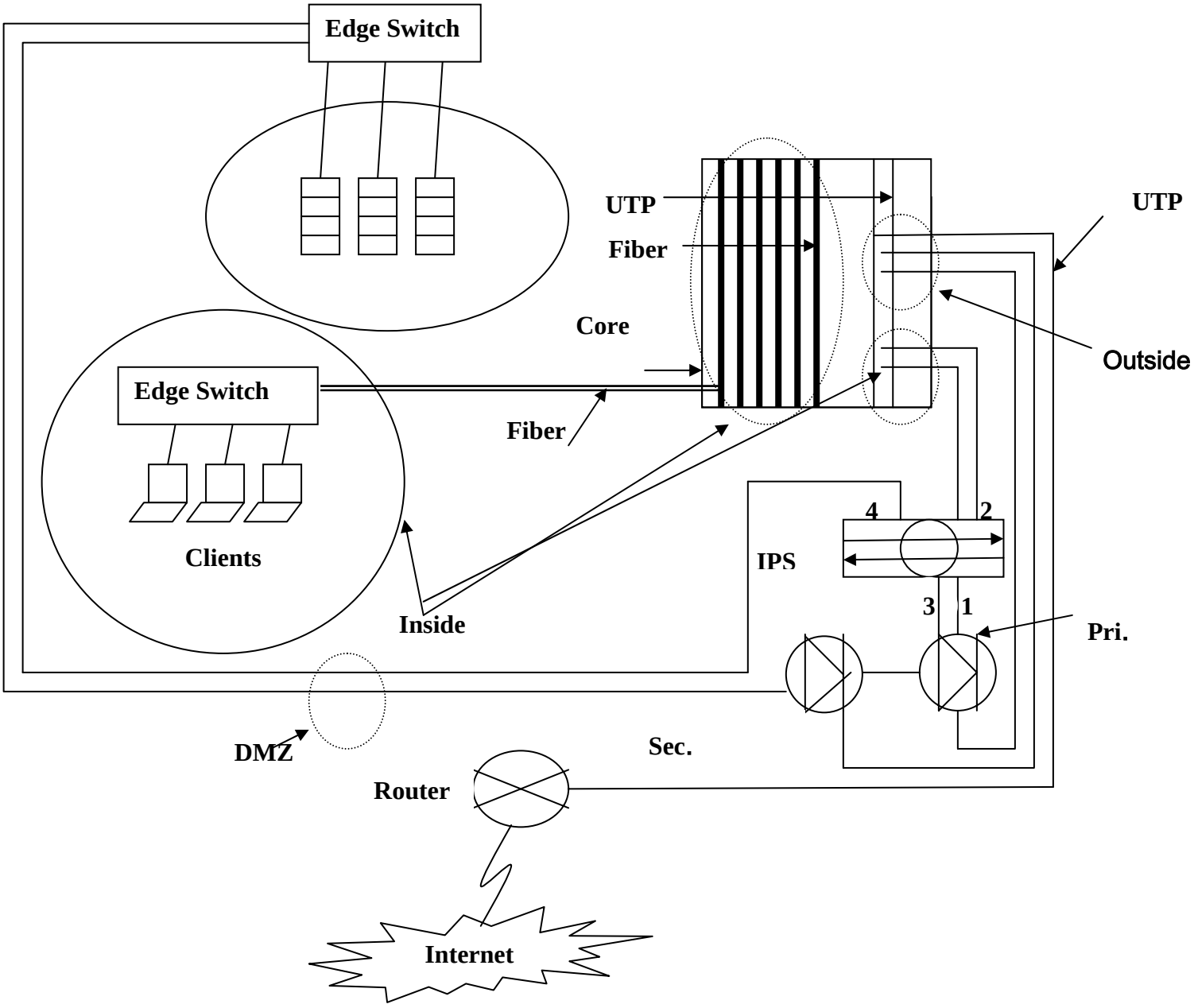

Figure 4: Security Layout for Dokki/Giza Campus Network

Figure 4 indicates the design of security system, which is based on isolating servers in separate zone and filtering traffic to and from them. Filtering traffic coming to the inside network.

2.2 Analysis for the security events that occurs on RADCON/VERCON System

In this section we will analyze all the events that have been occurred during 24 hours to indicate the situation for the traffic from security point of views. This analysis is based on logging of firewall that have been exported to my PC and converted to graphs and tables using Manage Engine Firewall Analyzer software:

2.2.1 Inbound /Outbound Traffic

2.2.1.1 Inbound Traffic

Figure 5 and table 8 indicate the amount of the inbound traffic through the firewall system which equals to 9834.1 MB during 24 hours this amount the traffic comes from Internet to the ARC network. Moreover, the hosts for the inbound traffic is determined in table 8 for example the host 192.100.32.4 used 1137.37 MB of the inbound traffic and this IP is belonged to the Agricultural Engineering Research Institute (AENRI) also we can see which services used by this host and from these services we can know if this traffic is normal or not; for this case especially I have found this host uses http service, Bittorrent, peoplesoft, ftp, edonky2000, pptp, and unknown services. Bittorrent: it is a peer to peer program, prior to version 3.2, Bittorrent by default uses ports in the range of 6881-6889. As of 3.2 and later, the range has been extended to 6881-6999. (These are all TCP ports, Bittorrent does not use UDP.) The client starts with the lowest port in the range and sequentially tries higher ports until it can find one to which it can bind. This means that the first client you open will bind to 6881, the next to 6882, etc. Tuxedo: is a middleware product that uses a message-based communications system to distribute applications across various operating system platforms and databases. The actions that will be taken for this host is: modifying firewall configuration to not accept Bittorrent traffic, examine this host for other services, and make some awareness for such users for the dangers of using this programs.

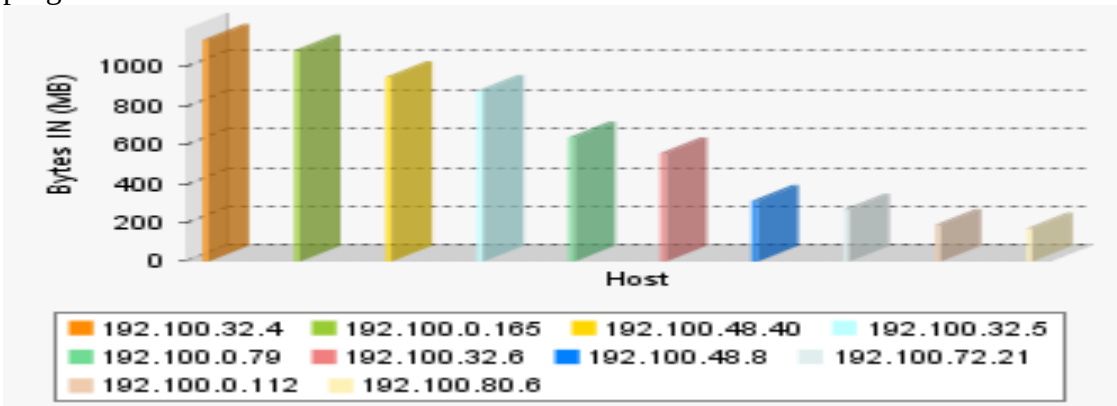

Figure 5: Inbound traffic

If we look at the second host we will find its IP is 192.100.0.165 which is belonged to for the Central Laboratory of Agricultural Expert Systems. The services that are used by this host are: HTTP, HTTPS and MSSQL services, and so on for the other hosts.

Table 8: Top hosts for inbound traffic

Host	Hits	Bytes In(MB)
192.100.32.4	180480	1137.37
192.100.0.165	3860	1083.31
192.100.48.40	163104	946.51
192.100.32.5	22186	876.73
192.100.0.79	8077	636.21
192.100.32.6	29629	555.45
192.100.48.8	3697	311.64
192.100.72.21	16211	265.68
192.100.0.112	3193	187.77
192.100.80.6	5716	169.11
Others	904230	3664.33
Total	1340383	9834.1

2.2.1.2 Outbound Traffic

Figure 6 and table 9 indicate the amount of the inbound traffic through the firewall system which equals to 202.57 MB during 24 hours this amount the traffic comes from the ARC network to Internet. Moreover, the hosts for the outbound traffic is determined in table 9 for example the host 192.100.24.17 which is belonged to Animal Production Research Institute (APRI) used 103.49 MB and if we analyze the used services by him we will find that he used services are: DNS, SMTP, HTTP, and Unknown service. From the analysis for the used service there is a big possibility that this host is infected by a virus that sent messages to many destination, therefore it should be examined to block this behavior. Another example is the host 192.168.1.12 which is the mail server of CLAES used 76.91 of the outbound traffic, from the analysis of the used service I have found that it is only SMTP, but the problem that I have found this machine sent messages in non working hours after 12 p.m. to some destinations a sample of the output file that illustrating this behaviors as the following:

Destination	**Protocol**	**Time**	**sent (KB)**
207.234.215.17	smtp	19th Apr 2007, 00:00:01	8.38
128.32.222.29	smtp	19th Apr 2007, 00:00:01	2.65
144.140.80.13	smtp	19th Apr 2007, 00:00:01	0.5
155.181.144.21	smtp	19th Apr 2007, 00:00:01	0.62
155.181.144.22	smtp	19th Apr 2007, 00:00:01	0.62
155.181.144.23	smtp	19th Apr 2007, 00:00:01	0.62
155.181.177.26	smtp	19th Apr 2007, 00:00:01	0.62
155.181.177.27	smtp	19th Apr 2007, 00:00:01	0.62
155.181.177.28	smtp	19th Apr 2007, 00:00:01	0.62
192.100.122.233	smtp	19th Apr 2007, 00:00:01	1.34
195.128.174.72	smtp	19th Apr 2007, 00:00:01	3.62

As this is a very serious and fatal problem for CLAES Mail server, we are currently exert efforts to fix this problem by moving the current mail server (MS Exchange

2000) to the higher version (Exchange 2007) to use its spam engine, in addition to make sure that the new machine is not compromised by any mail hostile programs.

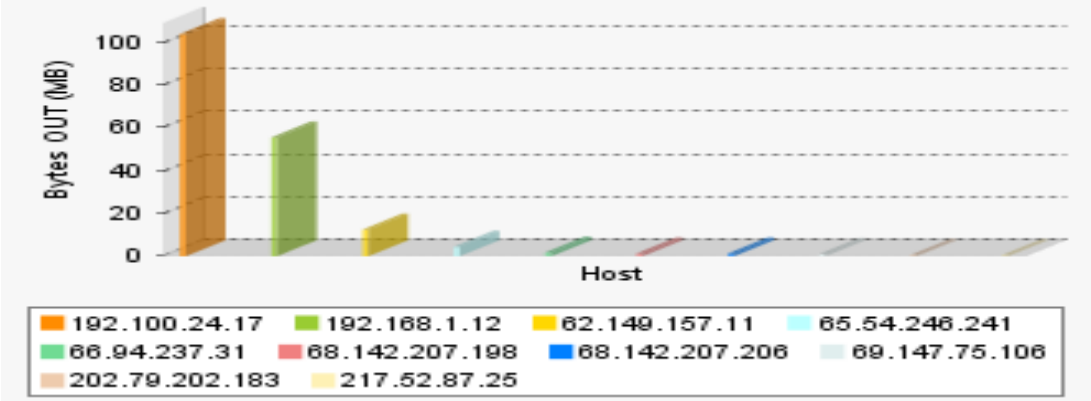

Figure 6: Outbound traffic

Table 9: Top hosts for outbound traffic

Host	Hits	Bytes Out(MB)
192.100.24.17	43455	103.49
192.168.1.12	17904	76.91
62.149.157.12	2	44.96
132.229.12.150	5	19.9
62.149.157.11	1	12.33
192.100.88.55	14382	8.77
192.100.0.253	642	5.17
65.54.246.241	2	4.43
66.94.237.31	2	1.72
192.100.0.149	3542	1.4
Others	1260446	29.68
Total	1340383	308.76

2.2.2 Protocols Traffic Reports

2.2.2.1 Top Protocols that sent traffic

Figure 7 and table 10 indicate the volume of sent traffic used based on the type of protocol. From the analysis for the result we note that mail protocols are the most used protocols and they sent 328.61 MB in the specified period by using SMTP. The top hosts that use SMTP to send messages are: 192.168.1.12 which is the mail server itself (46.96 MB), 192.100.24.17 (31.4 MB) and so on. The other protocols sent very small traffic with respect to SMTP server as indicated in table 10.

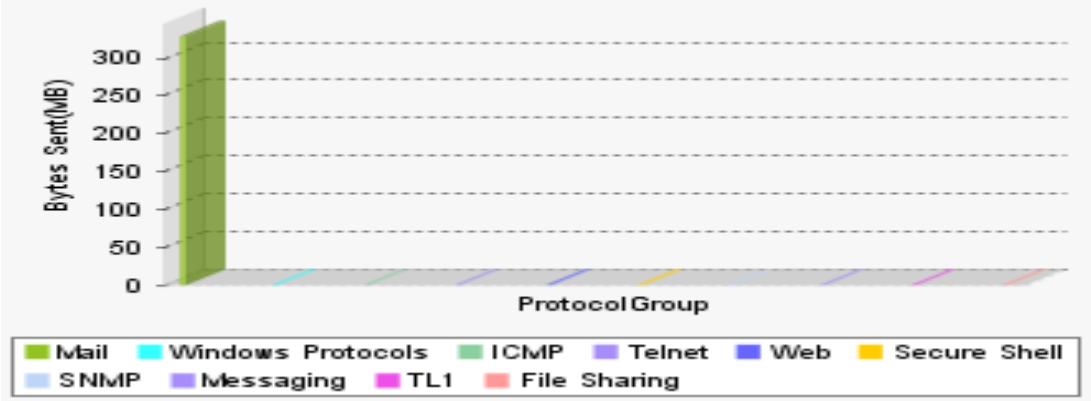

Figure 7: Top Protocols that sent traffic

Table 10: Top Protocols that sent traffic

Protocol Group	Hits	% Hits	Bytes Sent(MB)	% Bytes Sent
Mail	39700	1.65	328.61	100
Windows Protocols	82084	3.42	0	0
ICMP	40446	1.68	0	0
Telnet	29	0	0	0
Web	185465	7.72	0	0
Secure Shell	942	0.04	0	0
SNMP	83	0	0	0
Messaging	3359	0.14	0	0
TL1	41	0	0	0
File Sharing	9317	0.39	0	0
Others	2041363	84.96	0	0
Total	2402829	100	328.61	100

2.2.2.2 Top Protocols that received traffic

Figure 8 and table 11 indicate the volume of received traffic used based on the type of protocol. From the analysis for the result we note that Web protocols is the most used received protocols and it uses 5541.22 MB followed by Unassigned protocols that consumes 3407.71 MB and database applications that consumes 659.44 MB. The top hosts that received web traffic are 192.100.0.165 (1067.12 MB), 192.100.48.8 (301.36 MB), and 192.100.0.112 (175.88 MB). The assigned protocols are unknown; Bittorrent, tuxedo, and acmsoda and they use 34.58 percent of the received traffic. The top hosts that received unassigned traffic are 192.100.32.4, 192.100.48.40, 192.100.32.5, they are identified and they will be examined

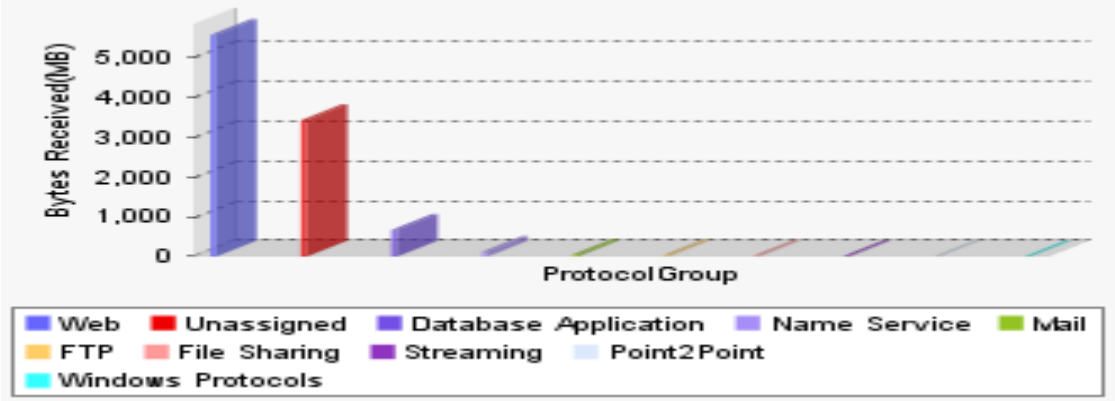

Figure 8: Top Protocols that received traffic

Table 11: Top Protocols that received traffic

Protocol Group	Hits	% Hits	Bytes Rcvd (MB)	% Bytes Rcvd
Web	185465	7.72	5541.22	56.24
Unassigned	1661107	69.13	3407.71	34.58
Database Application	91619	3.81	659.44	6.69
Name Service	285434	11.88	118.71	1.2
Mail	39700	1.65	44.8	0.45
FTP	768	0.03	26.1	0.26
File Sharing	9317	0.39	25.54	0.26
Streaming	654	0.03	15.88	0.16
Point2Point	77	0	10.36	0.11
Windows Protocols	82084	3.42	3.36	0.03
Others	46604	1.94	0.45	0
Total	2402829	100	9853.58	100

2.2.3 Web Usage Traffic Reports

2.2.3.1 Top Hosts that sent Web traffic

As we have found that Web traffic is consuming a high volume of traffic, we have to make more analysis to this type of traffic. Figure 9 and table 12 depicts the volume of the sent web traffic related to which hosts. The abnormal entry here is the first host 222.126.20.82 is not related to ARC network so how it can sent traffic so, IP spoofing should be analyzed and if there is any necessary modifications for security devices they should be applied

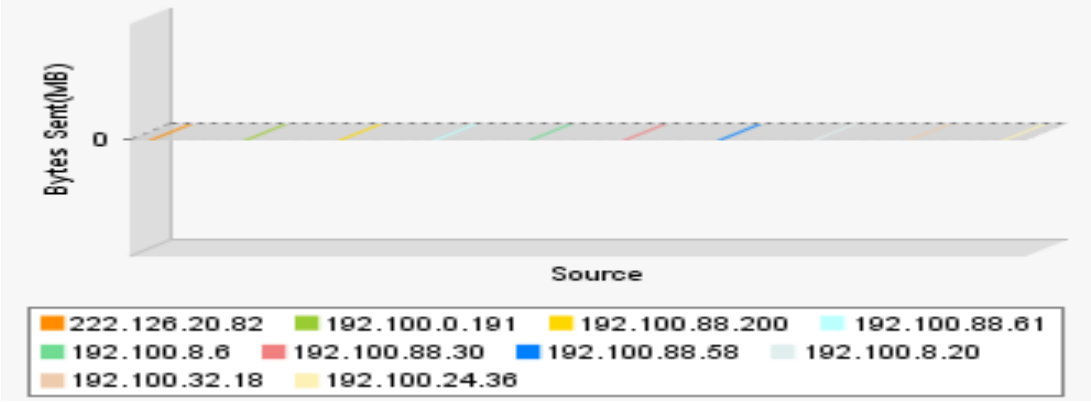

Figure 9: Top Hosts that sent Web traffic

Table 12: Top Hosts that sent Web traffic

Host	Hits	% Hits	Bytes Sent(MB)	% Bytes Sent
222.126.20.82	4	0.01	0	-
192.100.0.191	1740	3.14	0	-
192.100.88.200	176	0.32	0	-
192.100.88.61	295	0.53	0	-
192.100.8.6	63	0.11	0	-
192.100.88.30	33	0.06	0	-
192.100.88.58	174	0.31	0	-
192.100.8.20	120	0.22	0	-
192.100.32.18	652	1.18	0	-
192.100.24.36	849	1.53	0	-
Others	51254	92.58	0	100
Total	55360	100	0	100

2.2.3.2 Top Hosts that received Web traffic

Figure 10 and table 13 depicts the volume of the sent web traffic related to which hosts. It has been noticed that host 192.100.0.165 has received a very high volume of web traffic this should be checked to see if this normal or abnormal traffic.

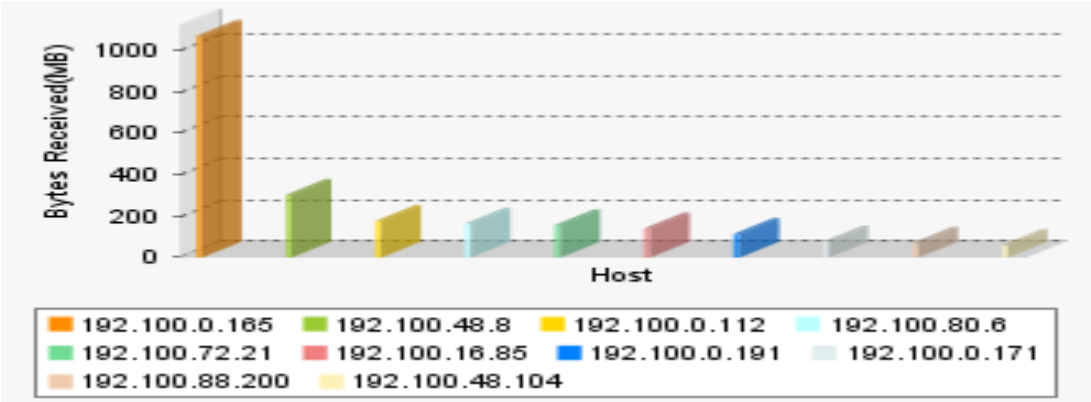

Figure 10: Top Hosts that received Web traffic

Table 13: Top Hosts that received Web traffic

Host	Hits	% Hits	Bytes Rcvd (MB)	% Bytes Rcvd
192.100.0.165	393	0.71	1067.12	22.87
192.100.48.8	337	0.61	301.36	6.46
192.100.0.112	855	1.54	175.88	3.77
192.100.80.6	863	1.56	162.97	3.49
192.100.72.21	2994	5.41	155.27	3.33
192.100.16.85	2284	4.13	139.15	2.98
192.100.0.191	1740	3.14	112.65	2.41
192.100.0.171	118	0.21	77.08	1.65
192.100.88.200	176	0.32	70.44	1.51
192.100.48.104	465	0.84	60.08	1.29
Others	45135	81.53	2344.72	50.24
Total	55360	100	4666.73	100

2.2.4 Mail Usage Traffic Reports

2.2.4.1 Top Hosts that sent Mail traffic

Figure 11 and table 14 depicts the top host that sent mail traffic, it has been noticed that mail server 192.168.2.12 sent 46.96 MB; other hosts sent very low traffic with respect to the first host. Configuration of Mail server should be reviewed fro how it handles relaying for outside IP addresses.

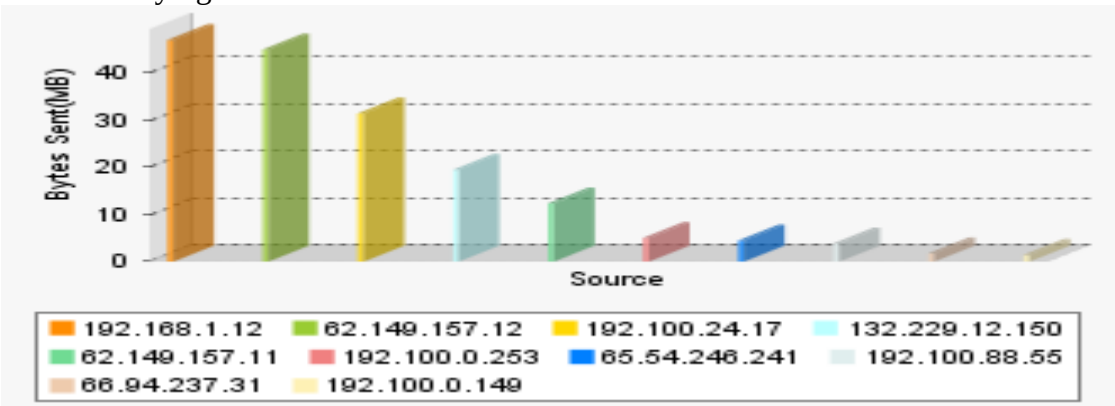

Figure 11: Top Hosts that sent Mail traffic

Table 14: Top Hosts that sent Mail traffic

Host	Hits	% Hits	Bytes Sent(MB)	% Bytes Sent
192.168.1.12	942	29.36	46.96	25.13
62.149.157.12	2	0.06	44.96	24.06
192.100.24.17	1379	42.97	31.4	16.81
132.229.12.150	2	0.06	19.42	10.39
62.149.157.11	1	0.03	12.33	6.6
192.100.0.253	1	0.03	5	2.68
65.54.246.241	2	0.06	4.43	2.37
192.100.88.55	428	13.34	3.79	2.03
66.94.237.31	2	0.06	1.72	0.92
192.100.0.149	2	0.06	1.4	0.75
Others	448	13.96	15.42	8.26
Total	3209	100	186.84	100

2.2.4.2 Top Hosts that received Mail traffic

Figure 12 and table 15 depicts the top host that received mail traffic, it has been noticed that host 19.100.0.165 received 12.12 MB and so on.

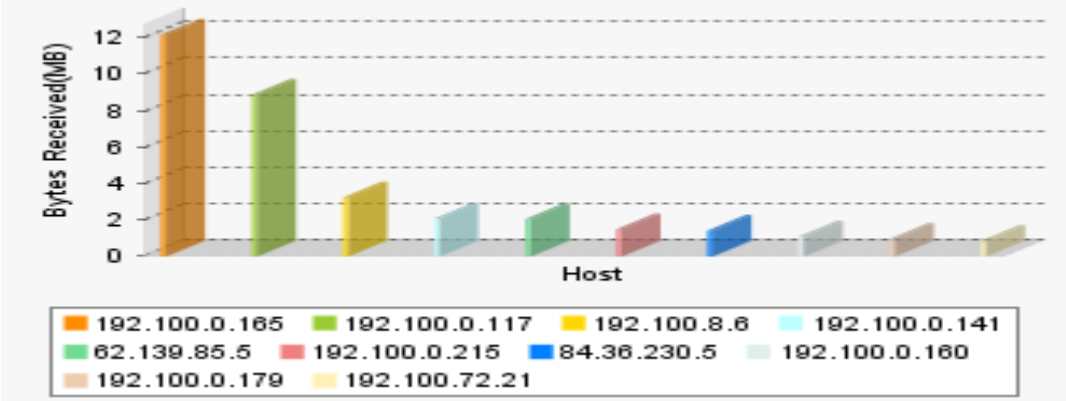

Figure 12: Top Hosts that received Mail traffic

Table 15: Top Hosts that received Mail traffic

Host	Hits	% Hits	Bytes Rcvd (MB)	% Bytes Rcvd
192.100.0.165	9	0.28	12.12	**30.48**
192.100.0.117	1	0.03	8.84	**22.24**
192.100.8.6	1	0.03	3.22	**8.09**
192.100.0.141	35	1.09	2.07	**5.2**
62.139.85.5	57	1.78	2.04	**5.12**
192.100.0.215	1	0.03	1.47	**3.71**
84.36.230.5	1	0.03	1.41	**3.55**
192.100.0.160	3	0.09	1.08	**2.72**
192.100.0.179	1	0.03	0.95	**2.39**
192.100.72.21	4	0.12	0.84	**2.1**
Others	3096	96.48	5.72	**14.4**
Total	**3209**	**100**	**39.76**	**100**

2.3 IPS Log Analyses

There is also in addition to the firewall analysis we can look at the IPS monitoring tool, the following that I have found during the same 24 hours. Figure 13 describe that there are events called SYN Host Sweep, and MSSQL Resolution Service Stack Overflow this is types of attacks. The action that has been taken is the modifications of the first event and trying to fix the second event.

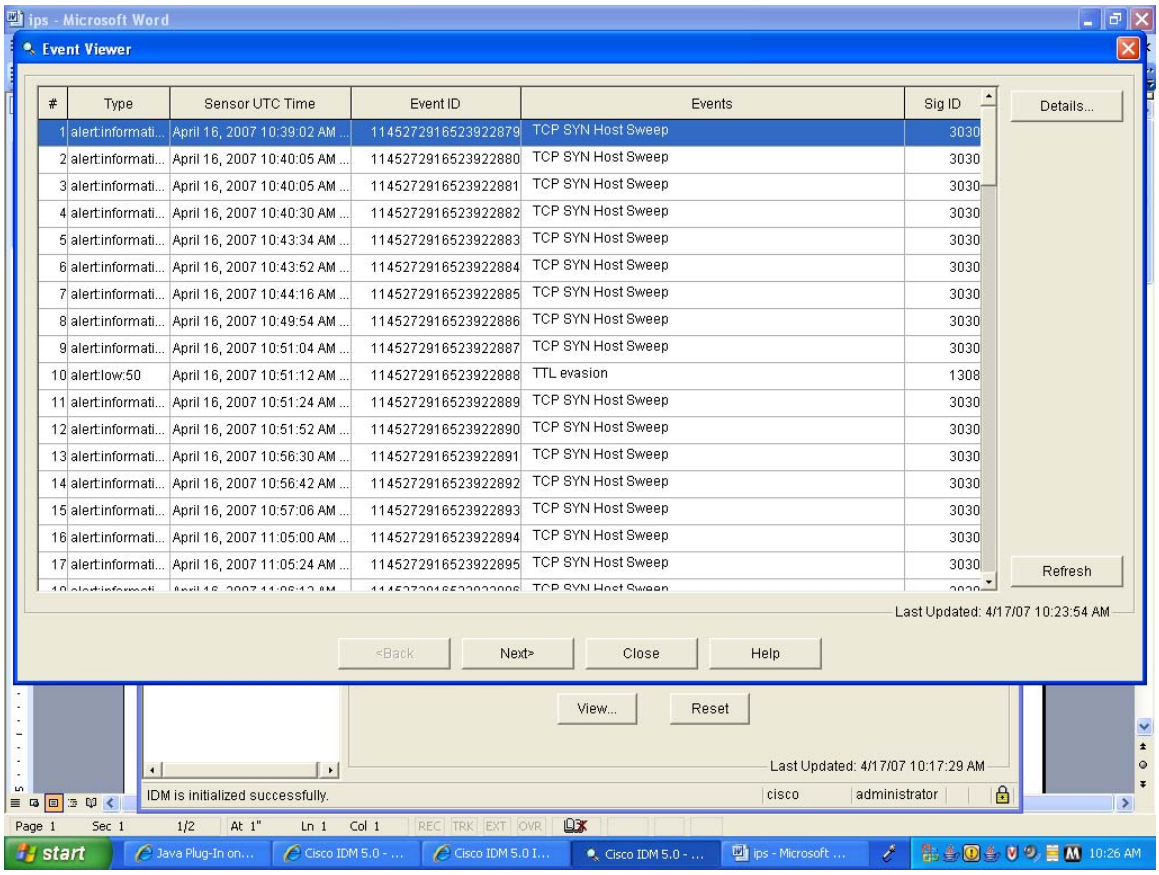

Figure 13: IPS event monitoring

The details of the first event were:

evIdsAlert: eventId=1145272916523922879 vendor=Cisco severity=informational
originator:
hostId: ARC-Dokki
appName: sensorApp
appInstanceId: 326
time: April 16, 2007 10:39:02 AM UTC offset=0 timeZone=UTC
signature: description=TCP SYN Host Sweep id=3030 version=S2
subsigId: 0
interfaceGroup:
vlan: 0
participants:
attacker:
addr: 192.100.0.130 locality=OUT
port: 3482
target:

addr: 97.135.81.195 locality=OUT
target:
addr: 97.135.81.193 locality=OUT
target:
addr: 97.135.81.199 locality=OUT
target:
addr: 97.135.81.197 locality=OUT
target:
addr: 97.135.81.203 locality=OUT
target:
riskRatingValue: 21
interface: ge0_1
protocol: tcp

The details of the second event are:

evIdsAlert: eventId=1145272916523922978 vendor=Cisco severity=high
originator:
hostId: ARC-Dokki
appName: sensorApp
appInstanceId: 326
time: April 16, 2007 6:43:41 PM UTC offset=0 timeZone=UTC
signature: description=MSSQL Resolution Service Stack Overflow id=4701 version=S137
subsigId: 0
sigDetails: MSSQL Resolution Service Stack Overflow
interfaceGroup:
vlan: 0
participants:
attacker:
addr: 209.208.170.226 locality=OUT
port: 1301
target:
addr: 192.100.0.52 locality=OUT
port: 1434
riskRatingValue: 85
interface: ge0_1
protocol: udp

2.4 Conclusion

From the previous mentioned analysis we conclude the following:

- Monitoring of security events is a continuous process but a methodology for this monitoring should be adapted including:
    1. Documentation of the current network resources, this has been done in the first progress report that includes a full description for the ARC network.
    2. Identifying the current protection controls including hardware and software equipment: firewall, IPS, patches, antivirus, software firewall, and anti spam engines
    3. Periodical assessment for the situation of the security, three months is proposed to be the interval for assessment, modifications, and reassessment

4. Study the impaction of enhancing security on network performance especially on servers, and Internet bandwidth.

- Modifications for security polices and adding other devices if needed should be taken according to the analysis report

3. Design and monitoring the implementation of a data protection and recovery plan for RADCON Servers.

3.1 Specifications of RADCON Servers and backup storage

3.1.1 Hardware

- Quantity: 3
- Dell, Power Edge 2800 – Xeon 3.2 GHz/1MB
- Memory: 2 GB DDR
- Two Hard Drives: 36 GB SCSI Ultra320 (15000 rpm)
- Three Hard Drives: 73 GB SCSI Ultra320 (10000 rpm)
- Optical Drive: 24 x CDROM
- Monitor: 17" TCO99 Flat Panel
- UPS: APC Back-UPS RS1000VA
- 4 External USB HDD, 250 GB

3.1.2 Software

- Operating System: MS Windows 2003 Enterprise Edition with service pack 1
- DBMS: MS SQL Server
- Web Server: IIS 6.0

3.2 Data protection plan

Before discussing the data protection plan we have to define our IT environment capabilities. There are basically three types of environments:

- **Direct Attached Storage (DAS) Environment:** DAS is the simplest backup and restore environment, usually consisting of a standalone backup drive (may be a tape or an external Hard Drive) attached directly to the server.
- **LAN Environment:** Storage backup devices are connected to the LAN and managed centrally from a single console through a single backup server reducing hardware costs and management time.
- **Storage Area Network (SAN) Environment:** Organization that runs SAN has similar characteristics to those that operate a LAN. But in addition they are likely to have a large and possibly complex network, needing as close to 100 percent uptime as possible.

RADCON Network has DAS and LAN data protection capabilities. The types of data that needed to be protected and the approaches for achieving protection for them:

1. MS Windows 2003 Enterprise Edition

There are basically two solutions to protect the operating system:
Redundancy of servers

In this case we buy for each application two server and put the same software on both of them; also we make the same modifications manually on the two servers. This method increase the time needed to recover the system because of the updating process and there should be a backup mechanism to allow recovery for the failed server. This method can be done with any operating system.

Clustering

There two technologies that are supported by Windows 2003: server clustering and network load balance (NLB). Server cluster is used for databases, e-mail services, and the most similar applications; can be deployed on a single network or geographically distributed; support up eight nodes; and it requires the use of shared storage. Figure 14 illustrates an example for server clustering.

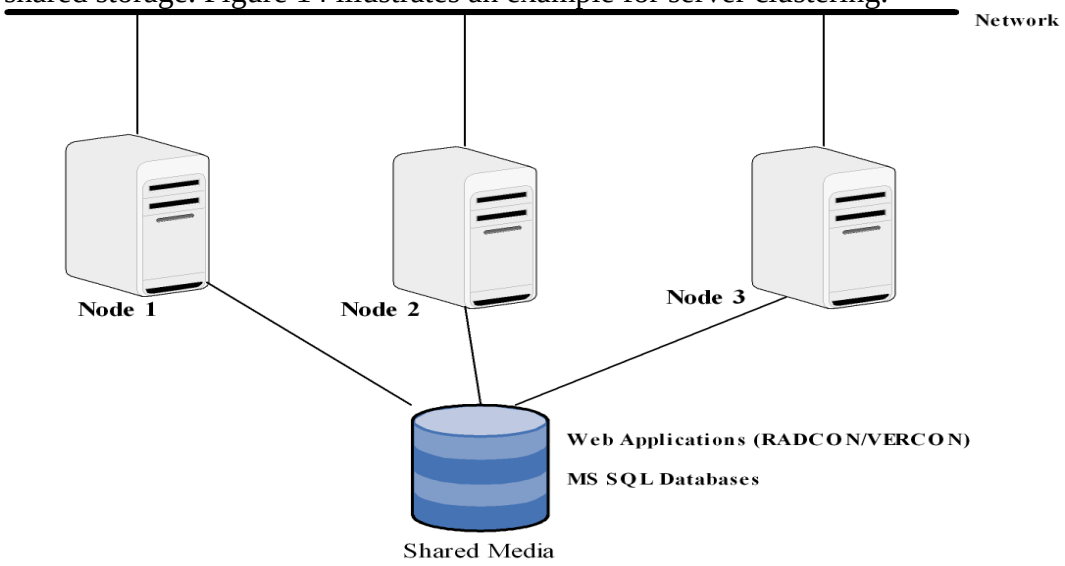

Figure 14: RADCON Servers Cluster

Connecting all nodes to a single storage device simplifies the challenge of transferring control of the data to a backup node. However, this architecture has weaknesses. If the storage device fails, the entire cluster fails. To solve this problem Majority node set (MNS) server clusters is used to store the quorum on a locally attached storage device connected directly to each of the cluster nodes.

NLB is used for Web serves, firewalls, and the most similar applications; usually deployed in single network; support up to 32 nodes; and it does not require any special hardware or software. Figure 15 illustrate an example for using NBL or RADCON system

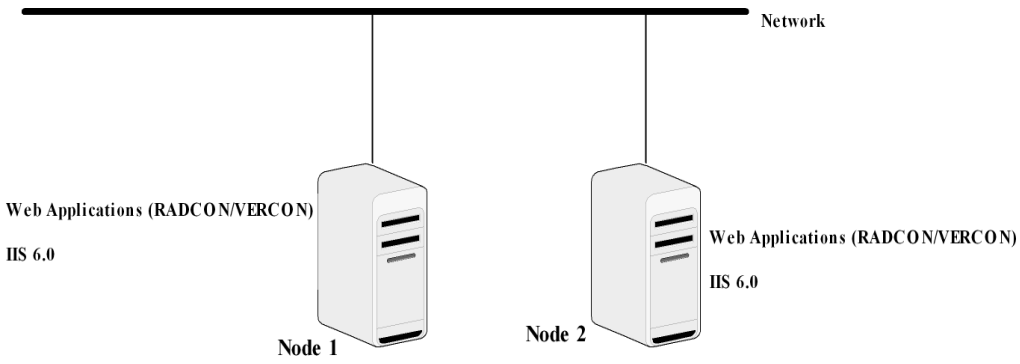

Figure 15:Protection of RADCON Servers using NBL

Both Server Cluster and NLB can be used to provide availability in the event of a failure of a processor, memory chip, power supply, or other hardware component. To provide complete redundancy, all layers of applications must be clustered.

The optimal design for RADCON system will be based in using 2 nodes contain Web Applications with NBL clustering and 2 nodes contain SQL Databases as shown in figure 16

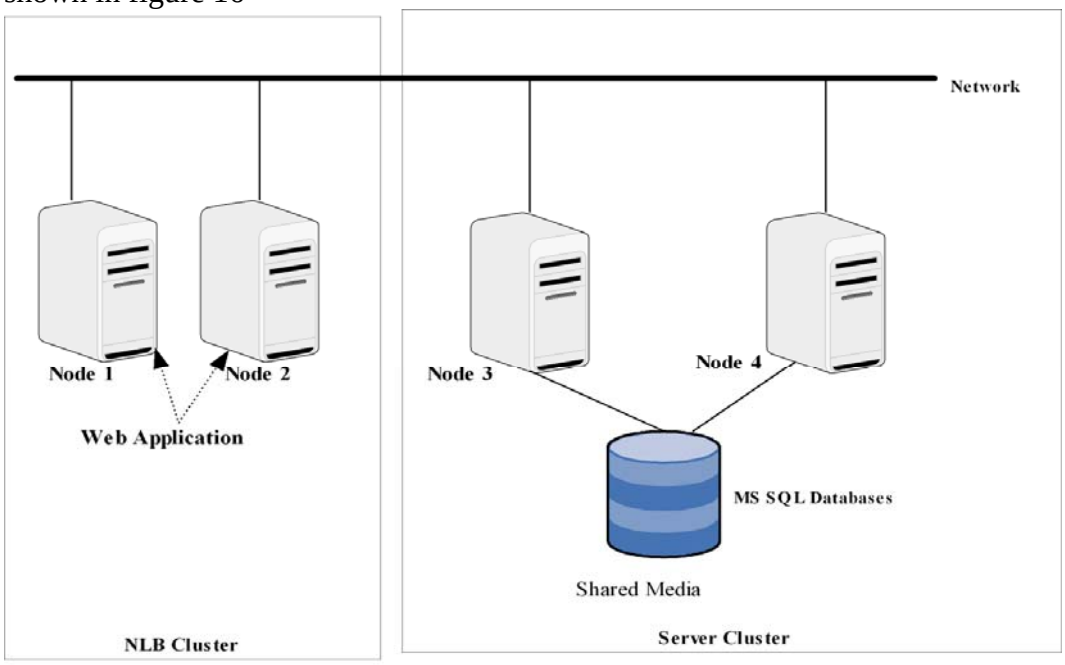

Figure 16: Optimal Design for using clustering for RADCON system

2. MS SQL Server 2000 Enterprise Edition

Correcting the SQL Server 2000 availability problem is difficult and therefore a number of technologies have been developed to assist organizations to meet their needs. These technologies include:

- **Replication**

SQL Server 2000 includes three types of replication: transactional, merge and snapshot, offering the option to replicate data from the publisher to one or many subscribers via a distributor to manage the process.

Transactional: Changes made to a publication by the publisher are replicated to a distributor and then on to a subscriber immediately. Transactional replication

is appropriate for situations where data in remote systems needs to reflect changes in a “master” database.

Merge: Similar to Transactional replication, but allows changes to be made in multiple places, either at the publisher or subscriber. Merge replication is appropriate for a situation where the data resides in multiple places for performance reasons.

Snapshot: Sends data as it existed at a specific point in time, regardless of any updates being applied to the data. Snapshot replication is appropriate for situations where you do not need up-to-the minute access to your data.

- **Log Shipping**

Log shipping is automatically copying and restoring the database's transaction logs to another database on a standby server as depicted in figure 17. Because the standby database receives all changes to the original database, it's an exact duplicate of the original database—out of date only by the delay in the copy-and-load process. After that we have the ability to make the standby server a new primary server if the original primary server becomes unavailable. When the original primary server becomes available again, we can make it a new standby server.

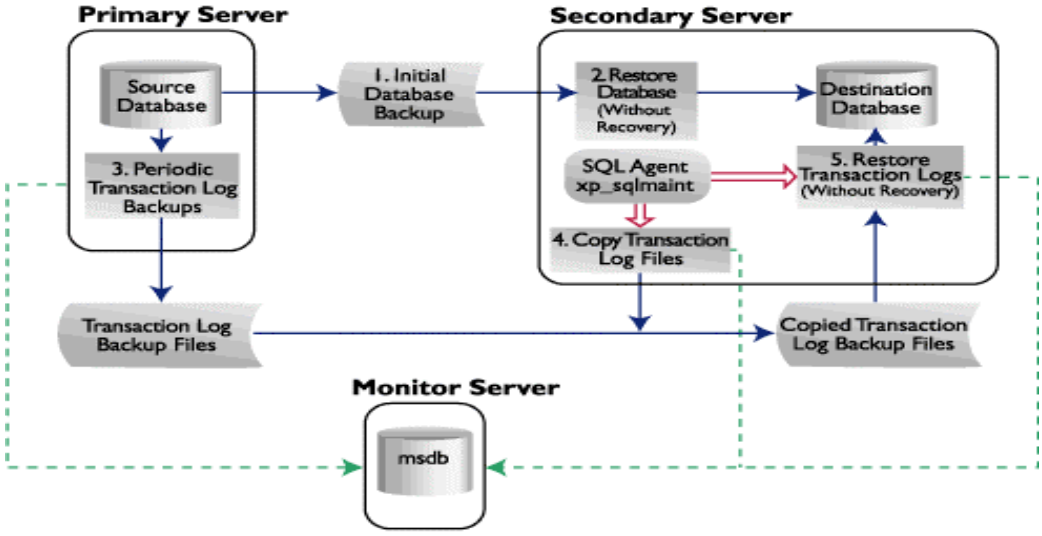

Figure 17: The operation of MS SQL Log Shipping

Log shipping is a type of high availability solution, and it works rather effectively. One of the biggest benefits of log shipping is that it is a much cheaper high availability solution than clustering. This is because the hardware requirements that are necessary for clustering are not required for log shipping.

- **Clustering**

SQL Server 2000 failover clustering is built on top of a Windows 2003 server as indicated in figure 18

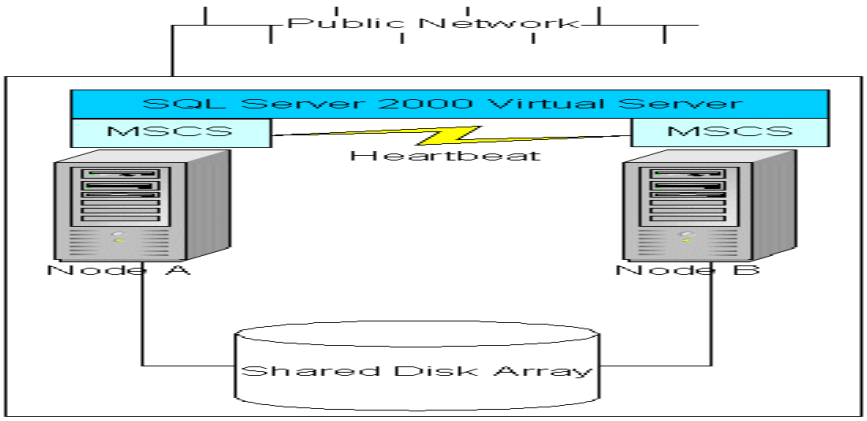

Figure: 18 MS SQL Server 2000 clustering

3. IIS 6.0

There are two resource to be completely protect IIS

1. to backup application files
2. to backup configuration files

3.3 Our Implementation Plan

The plan is based in using Microsoft solutions with the available resources and it is based on the following:

- Using MS Windows 2003 Enterprise Edition for the three servers
- Using One server as Web Application while the other will be used as Database Application
- The third server will operate as a backup for both Web and Database Applications
- A periodical backup will be made for Web and Database Applications using:
  - External HDD on weekly periods
  - Through network on light load periods on three days basis
- A replication between the main Database Application server and a Backup Server using transactional model will be made

There is some delay in this activity because of technical problems in the implementation of transactional replication and the verification of it. Also a scenario for recovery case will be adapted to ensure its operation