The 1st IEEE International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems
20-21 September 2012, Offenburg, Germany

A Highly Scalable Monitoring Tool for Wi-Fi Networks

Pheeha Machaka[1], Antoine Bagula[1] and Nico De Wet[2]
[1]Intelligent Systems and Advanced Telecommunication Laboratory (ISAT)
Department of Computer Science, Room 317 Computer Science Building, 18 University Avenue, University of Cape Town, Rondebosch, Cape Town, South Africa, 7701
and [2]RedButton Mobile CC, C/O Bandwidth Barn, 125 Buitengracht Street, Cape Town, 8001
email: [1]{pheeha.machaka, antoine.bagula}@uct.ac.za; [2]nico@redbutton.co.za

***Abstract* - The paper introduces a monitoring tool that was designed for an existing network of Wi-Fi hotspots. This was done by adding data collection and visualization components to the existing network. Syslog protocol was used for data collection and left running for two months monitoring network's performance. Google Maps was used for Visualizing overall network's performance. The tool was tested with experts and it was compared to existing monitoring tools and It was found that the monitoring tool was more scalable and effective than traditional monitoring tools.**

Keywords* - *Management; Measurement; Performance Monitoring; Middleware; Wi-Fi Networks; Service Level Agreement; Visualization

I. INTRODUCTION

Wireless Fidelity (Wi-Fi) is a wireless networking technology that uses radio waves to provide high-speed wireless internet and network connections. This technology is based on the IEEE 802.11 group of standards, including the 802.11a, 802.11b and 802.11g. This technology has proven to be a fast-wireless networking approach that is relatively easy and inexpensive to implement [1]. This is made possible by using wireless access points (AP) or hotspots. The AP broadcasts signals to Wi-Fi-capable devices within the AP's range. The AP's are deployed in different business settings, like coffee shops, restaurants, hotels and conference rooms and capable devices range from laptops to cellular mobile devices and Personal Digital Assistants (PDA).

To ensure optimal performance of the network, one needs to monitor its performance. The service provider agrees, in measurable terms the service level that the organization requires from the network, a Service Level Agreement (SLA). Network administrators have previously relied on a reactive approach to monitor networks, but in the case of a large network, this may present service delivery problems.

- How can one develop a monitoring tool that can be scalable to a large Wi-Fi Network? What does this system need to include?

[1] The paper attempts to answer the questions and introduces an approach to monitoring networks by studying management of networks on a large scale. A monitoring tool was developed by adding to an existing one, three monitoring components, the data collection, visualization, storage and retrieval components.

II. RELATED WORK

In [2] a Monitoring tool (SEQUIN) for a Multiprotocol Label switching (MPLS)-based network was developed. A MPLS-based network allows packet switching at a high rate, while retaining the flexibility of Internet Protocol (IP). It enables Service Providers to preserve network features where IP provide support for Quality of Service (QoS) through differentiated services. To ensure that the level of QoS is met, service providers need to monitor the network and keep track of Bandwidth utilization and other QoS metrics. SEQUIN uses SNMP-based techniques to keep track of QoS metrics for a network service provider. This monitoring tool had the same modules as the one presented in this paper. The modules' different tasks included network polling, computation of QoS Metrics, and visualizations.

SEQUIN's database structure, classified data stored in the database as static or dynamic. Static information includes system configuration, network element configuration, network topology, and monitoring agent information. Dynamic information consists of polled SNMP data and information computed with this data. The database kept QoS statistics while polled data was cleared frequently, thus making the database more scalable.

In [3] the history of the Multi Router Traffic Grapher (MRTG) is described. The first version of the MRTG program was a Perl script which makes SNMP queries and creates images for display in HTML. The third version of MRTG used the Round Robin Database (RRD). With this storage mechanism, MRTG became faster and more configurable. MRTG has moved from simple, plain files' storage mechanism to a more sophisticated and reliable database storage mechanism.

The financial assistance of the National Research Foundation (NRF) and Telkom Centre of Excellence (CoE) is hereby acknowledged.

This shows that simple ASCII files can be used for network monitoring, but they have greater performance disadvantages [3].

Cacti [8] is a front-end for RRDTool. It stores all the necessary information to create graphs in a MySQL database. The front-end is written in PHP. Cacti does the work of maintaining graphs, data sources, and handles the actual data gathering. There is support for SNMP devices, and custom scripts can easily be written to poll virtually any conceivable network event.

Nagios [9] is an open source program that monitors hosts and services on the network. It uses SMP for real-time polling on devices for network health indicators, and it uses and SQL databse for data storage. It also offers real-time event notification via sms or email to the relevant person.

SmokePing [10] is a deluxe measurement tool that measures, stores and display latency, latency distribution and packet loss all on a single graph. SmokePing uses RRDTool for data storage, and can draw very informative graphs.

These monitoring tools will later be compare to the monitoring tool developed for the paper and the next section of the paper will discuss the overview and motivation of designing a monitoring tool for a network of Wi-Fi hotspots.

III. THE WI-FI HOTSPOT SYSTEM

The monitoring tool designed in this paper was designed for a company that deployed a large network of hotspots, RedButton CC. They have a network of more than 400 hotspots with more than 615 gateway devices interconnected. The section discusses the old network and the new network that the new monitoring tool will introduce.

A. The Old System

In a large network of wireless devices, it becomes difficult for a network administrator to monitor and manage the network without using proper monitoring tools that don't have numerical and graphical representation. This may lead to a reactive response to managing the network. This is the case with Redbutton CC, our case study network.

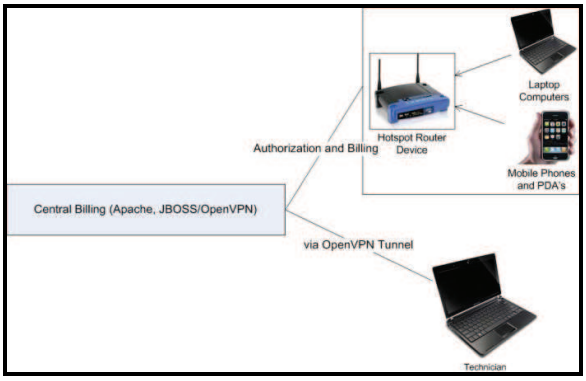

Figure 1. Current Situation for the Network

Figure 1 shows the setting for the current situation for the case study network. In this case, any Wi-Fi capable device within the hotspot range can connect to the network and access the world wide web. The user device will connect to the router, and this router will connect to a central billing server for payments and billing. The technician does not have a way of monitoring the performance of the network in this case.

B. The New System

The existing systems studied in section 2 had methods for performing data collection and visualization of the data. Therefore, the monitoring system designed in this paper introduces three new components to the network. The data collection, storage and retrieval, and the visualization component. This is depicted in figure 2 below.

The data collection component will collect performance data from each device that is connected to the network (the details will be further discussed in section 5). The data collected from each device is stored in a central storage and retrieval component, where the data will be summarized using data mining techniques.

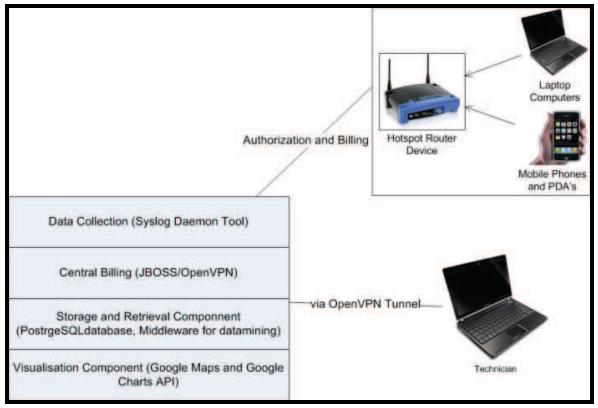

Figure 2. New network with more components

The visualization component will display the performance of the network in a way that the network administrator will understand and make good decision (the details of the visualization component will be discussed in section 6).

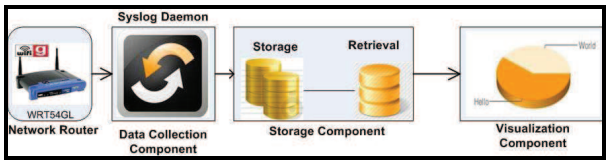

Figure 3. Indicating the flow of information between the three new components

Figure 3 indicates how the three new components will communicate to each other and how information will flow from one to the other.

The next section of the paper will discuss the various performance metrics that will be monitored and displayed by the system.

IV. PERFORMANCE METRICS

We conducted a set of experiments based on three performance metrics which are usually used in performance evaluation of Wi-Fi networks. These are:

Uptime and Downtime. This metric measure the time a device has been up and running. It reveals the availability, stability and reliability of the communication device.

Load Average. Measures the "congestion rate" for the device based on the number of users connected to the device.

Radio Noise and Channel. Wi-Fi uses the 2.4 GHz spectrum band which is shared with other devices like cell phones, GPS, RFID tags and Bluetooth devices. Note that the proliferation of devices using the free 2.4 GHz ISM band leads to more congested and noisy Wi-Fi devices.

V. DATA COLLECTION

For the data collection component, two data collection methods were investigated, Simple Network Management Protocol (SNMP) and Syslog Protocol. The section will discuss the message sent over the network and bandwidth consumed per message.

A. Simple Network Management Protocol

SNMP is non-proprietary and commonly used by network administrators. With SNMP one management system can communicate with devices from multiple vendors. All SNMP-enabled devices contain a specific text-file called the Management Information Base (MIB), which is a collection of hierarchically organized information that defines what data can be collected from that particular device using the protocol. The SNMP management station queries the devices(client) using commands to obtain device-specific information. It uses a client-server mode of communication, where both the client device and the manager device will send messages towards each other.

1) SNMP Message Overheads

The variable binding feature of the SNMP message allows the SNMP manager to request more than one parameter reading per message [4]. In this case, the manager will request four performance metrics from the client, while the client responds with the information requested. It was found that the SNMP message in both directions of communication will use 206 bytes per message.

The SNMP manager will collect the performance data from the clients every hour.

$$31\ days \times 206 \frac{bytes}{hour} \times 24\ hours = 153264\ bytes$$

$$and\ 153264\ bytes \approx 0.14\ MB\ per\ month\ per\ device$$

The section that follows will now discuss the Syslog data collection method.

B. Syslog

Syslog **(RFC3164)** is also a client-server protocol that provides a framework under which machines (agents) can send event notification messages across an IP networks to event message collectors - also known as Syslog Servers or Syslog Daemons. Initially Syslog messages are stored locally; and these messages will be automatically routed to a central location. These messages are received by the logging host; the logging host has significant disk storage for incoming messages (stored in a database) [6].

The next section will now discuss the Syslog message.

1) Syslog Message Overheads

With Syslog, the daemon tool is installed on the client device, and therefore, the client will only communicate with the manager, not the other way around. This is a one-way communication mode. The data is stored locally, then transported to the manager at a set time.

It was calculated that a Syslog message with the four parameters discussed in section 4 will consume 142bytes.

$$31\ days \times 142 \frac{bytes}{hour} \times 24\ hours = 105648\ bytes$$

$$and\ 105648\ bytes \approx 0.10\ MB\ per\ month\ per\ device$$

From this investigation, SNMP uses 40% more bandwidth than Syslog. The long run costs of using SNMP are greater.

The Syslog protocol was ran from 2009-07-03 to 2009-09-02 and 356537 records of data were collected.

VI. VISUALIZATION

The paper presents a case study network that covers a large geographical area. The hotspots are placed in strategic business areas across the city. The network has grown to accommodating more than 400 hotspots across the city. Google maps was used to visualize the overall graphical performance of the network, while Google Charts was used for numerical representations of network performance.

The next subsections will give details of how visualization for the tool was carried out.

A. Google Maps

Google Maps has an Application Programming Interface (API) for developers available. Graphical visualization of the monitoring tool was arranged in a hierarchical order. Figure 4 displays the order in which visualization was arranged. In the first level of graphical representations, the hotspots will be arranged into groups according to their geographical location. This makes it easy to summarize performance by location. The second level will now focus on performance representations for the hotspots, while the third will focus on the router devices in a chosen hotspot.

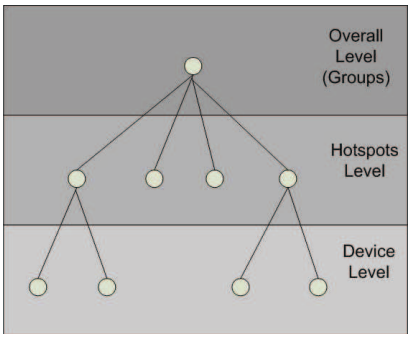

Figure 4. Structure of Visualisation

Figure 5 shows a snapshot of the monitoring tool's map representations. The main interface is divided into the maps section and the summary statistics section.

The main interface will display the performance of the network using color codes (Green, Yellow and Red). Displaying problematic areas or hotspot that need attention. The summary statistics section will give a numerical summary of the network performance at the level that the network administrator chose.

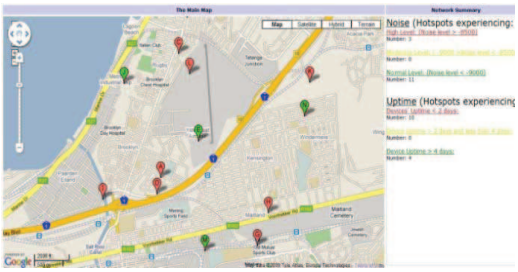

Figure 5. The main interface of the monitoring Tool

B. *Google Charts*

Google charts is also one of Google's products that is available to developers for free. This was used for historical performance reporting. The network administrator can choose to view the device or hotspot's historical performances, and they will be presented with graphs and charts displaying performance trends. Figure 6 is a snapshot of the uptime history of a device. It also uses color codes to indicate performance trends. The uptime history pie chart will display the total uptime, downtime and the current uptime for the device or hotpot.

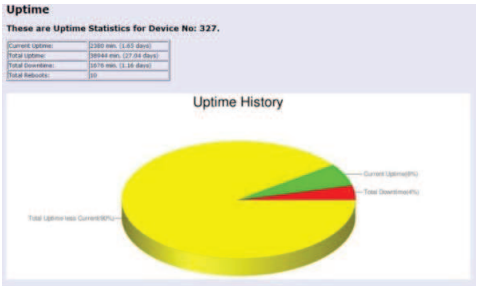

Figure 6. Device uptime history chart.

The next section of the paper will look into the evaluation of the monitoring tool.

VII. EVALUATION

The system was evaluated in terms of its functionality, visualization, ease of use, the usefulness and effectiveness of the information presented; and most importantly, scalability.

To do this, two ways of evaluating the system were considered; the monitoring tool was compared with existing systems and it was evaluated by experienced people who are working in the field of network monitoring.

A. *Comparison with Existing Systems*

This section of the evaluation process, the monitoring tool was compared with existing systems, namely **Cacti, Nagios and SmokePing**.

1) *Scalability*

Cacti presents data to an administrator in a list format as shown in figure 7, a snapshot of Cacti. The administrator can select a device from the list and view it's information and performance data. This becomes a problem when the network has a large number of devices operating in the network and they need to be monitored. The administrator will have to remember a device by name or device code. As the network grows and expands, the task that the network administrator will have to fulfill will also grow, leading to a reactive approach to network monitoring.

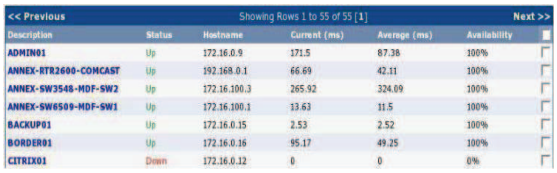

Figure 7. A snapshot of cacti's user interface, listing network devices (www.cacti.net)

With **Nagios**, they have a "network map" for visualizing the activities in the network, but these devices in the network are grouped by the domain into which they belong.

In this network map, they also display the status of the device, "UP" or "DOWN". This does not give the network administrator an indication of the full status of the device. This is shown by figure 8.

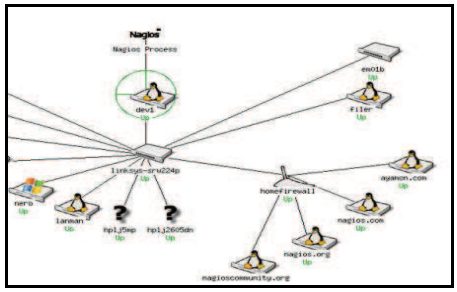

Figure 8. A snapshot of Nagios' visualization tool (www.nagios.org)

SmokePing is also another monitoring tool that uses RRDTool to visualize the network information. With SmokePing, they are also listing their network devices in the same way as Cacti. This way of representing network information is not scalable. It leads to problem once the network grows. Indeed RDDtool records data for only a certain amount of days, replacing previous data with newer data. This makes difficult to achieve current situation recognition and analysis based on historical data. It also makes difficult to the recognition of future situation based on current situations. A snapshot of how their data is represented, is shown in Figure 9.

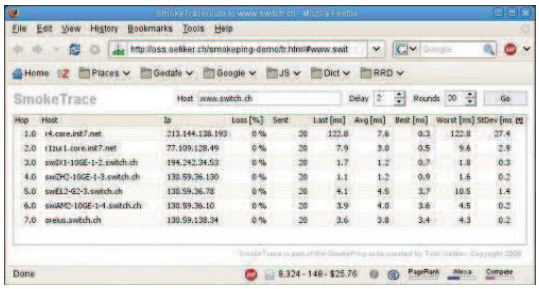

Figure 9. A snapshot of SmokePing's Monitoring Tool (http://oss.oetiker.ch/smokeping/)

The monitoring tool developed can group and position the devices in the network according to their geographical location, making it a more powerful and scalable visualization tool.

The next section will discuss the numerical and graphical representation abilities of these monitoring tools.

2) Graphs

From what was studied about Cacti, Nagios and SmokePing it was found that they produced customized graphing solutions to the network administrator, meaning that one can choose what they want to see on the graphs. This is a very helpful feature of the monitoring tool. Their graphs are in real time. An example of one of the graph generated by these monitoring tools is shown in figure 10.

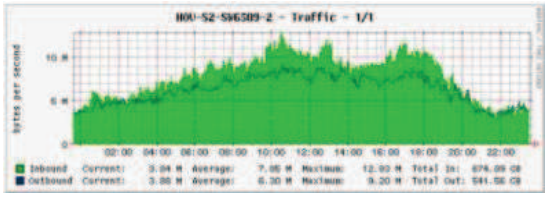

Figure 10. A snapshot of Cacti graphing tool

The graphs also provide summary information like the total, average, maximum and minimum readings for the metric on any router. These are some of the features that the monitoring tool had. This goes to show that the monitoring tool developed was on track with what network administrators would like to see on the graphs, to help them make decisions about the health of the network.

B. Expert User Feedback

For the expert user feedback section, an interview was conducted between the four team members in the company, and two computer science students with expertise in computer networking.

1) Visualization evaluation

In terms of scalability, the experts thought the monitoring tool was highly scalable, because with the monitoring tool, one can view network status at different levels (like: Device View, Hotspot View and Hotspot Group's view). They also thought that by grouping the hotspot by location, was a good idea, helping the administrator recall and find the device they want to view.

Using Google Maps makes it possible to visualize an area at many different scales. They also mentioned that users are more effective when using pictures of the real scenario, or problem they are tackling; therefore using pictures enables them to work more effectively and proactively. They also thought that with Google Maps service, the monitoring tool can be scalable to the level of visualizing the whole world, meaning there is no limit to the scalability.

One expert mentioned a very important scalability issue with the monitoring tool. They thought that the monitoring tool breaks down when monitoring hotspots that are on a high rising building. This is true because the hotspot will be in the same location, therefore the balloons on the main map will be on top of each other.

They also found the use of colors on the map very useful. The colors used to identify problem areas and areas that need attention. Using colors also helps in setting benchmarks for the key performance indicators (KPI). For example: if they want noise level on all routers to be below a certain level noise reading, they can then determine which routers are not falling within the target level, and those that need serious attention.

2) Graphs

In terms of the graphs generated by the system, the experts found them very helpful; and they liked the features of the tool, where one can view historical data and compare past events with current events, based on the data depicted by the monitoring tool's graph.

They also thought that the graphs can also help with measuring key performance indicators. For example: If the SLA stipulates that the device or the network should be running 99.9% of the time. They can easily measure this by looking at the graphs. The graphs presented in this tool show the uptime and the downtime of each device; therefore, one can tell if the SLA is met and if not, what can be the possible cause of problems. In this way, they can take steps into remedying the problem before the client can realize it or problems are encountered.

C. Data Collection Evaluation

For data collection, syslog data collection method was used to garner performance data from the devices connected to the network. This data can be used to design baseline or benchmark performance of the network. The KPI's can measure and verify critical services' availability, performance and scalability. This section looks at the metric performances of the network. In this case, the Performance Indicators for the technical point of view is discussed.

1) Noise Statistics

In a Wi-Fi device, the network administrator would like to have low level of noise for good performance. It would be preferable to have the network with a good performance level 99.999% of the times. Figure 11 indicates that 99.98% of the times, the network is in its good noise levels.

2) Load Performance

In a hotspot network, it would be favorable to have the traffic distributed equally among router devices. With a router's traffic load that is above 70%, the client may connect

but may be disconnected while in session; therefore, it would be preferred to have load performance of below 70%.

The load collected by the Syslog daemon programe is the 15 minutes average load for the device's load. It was found that 76% of the times, the network was experiencing close to 0 load, while 22.4% of the time the load was just below 10%.

This is indicative of a good performance of the network. Only 0.01% of the time, the network experienced a load that was above 100%.

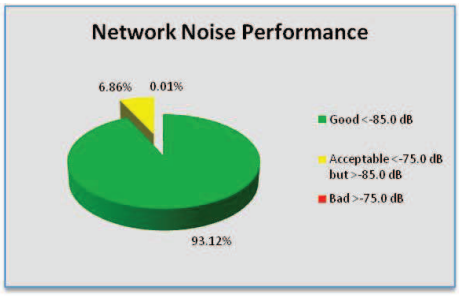

Figure 11. Network noise performance

This is indicative of a good performance of the network. Only 0.01% of the time, the network experienced a load that was above 100%.

With a load above 100%, the clients will connect but experiencing very slow connection and packets being dropped.

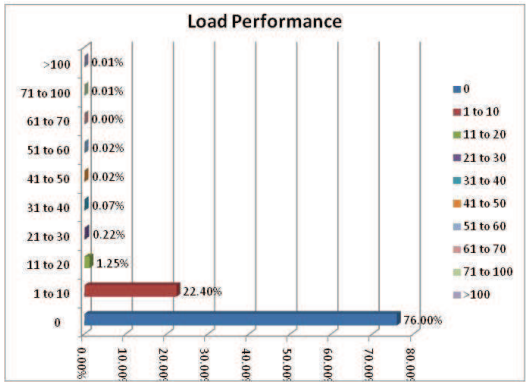

Figure 12. Network Load Performance

This is indicative of a good performance of the network. Only 0.01% of the time, the network experienced a load that was above 100%.

With a load above 100%, the clients will connect but experiencing very slow connection and packets being dropped.

3) Bandwidth Performance

Knowing how much bandwidth goes through each network device can help with capacity planning and management. The paper now reveals the network's bandwidth performance while considering the upstream and downstream performance.

TABLE I. DOWNSTREAM BANDWIDTH PERFORMANCE

Downstream Bandwidth (in Mb)	
Average	8.31
Minimum	0.00
Maximum	2561.92

From the data collected, on average, the data that goes through the network was 8.31 Mb per hour, while the maximum experienced was 2, 561.92 Mb per hour. Having a router device 2, 591.92 Mb per hour is an indication that there is a higher demand for more routers, thus growth for the network. This type of statistics that was garnered by the Data Collection component can help with capacity planning and management.

TABLE II. UPSTREAM BANDWIDTH PERFORMANCE

Upstream Bandwidth (in Mb)	
Average	3.20
Minimum	0.00
Maximum	1211.13

For the upstream bandwidth data, on average 3.2 Mb goes through the network in an hour, while the network experienced a maximum of 1, 211.13 Mb per hour.

This indicates that the network is experiencing consistencies in the amount of data that is downloaded from the network devices.

VIII. CONCLUSION

This system that was developed in this paper can help network administrators to proactively monitor a large Wi-Fi hotspot network's performance, without having to react when a problem occurs. Therefore improving the performance of the network and ensuring greater service delivery.

The statistics that was revealed in section 5.3 may be good, but not satisfactory to the network administrator. The administrator would like to have a more reliable network. In the future work of the monitoring tool, it is envisaged to find ways of detecting the faults or anomalies of the network using artificial immune systems and existing forecasting models.

REFERENCES

[1] Steven J. Vaughan –Nichols 2003. *The Challenge of Wi-Fi Roaming. IEEE Explorer*. July 2003.

[2] M. Thottan, G.K. Swanson, M. Cantone, T.K. Ho, Y. Ren, S. Paul, SEQUIN: An SNMP-based MPLS network monitoring system. *Bell Labs Technical Journal*, 2003, 95-111.

[3] Tobias Oetiker. 1998. MRTG: The Multi Router Traffic Grapher. In *Proc. of the 12th Conference on Systems Administration* (LISA '98). USENIX Association, Berkeley, CA, USA, 141-148.

[4] W. Stallings, *SNMP, SNMPv2 and CMI: The Practical Guide to Network Management Standards*. Addison Wesley. 1993.

[5] Donalds Pitts. Log Consolidation with Syslog. *GIAC practical repository,* SANS Institute, December 2003. [http://www.syslog.org/wiki/uploads/Main/log_consolidation_with_syslog.pdf]

[6] J. Riddel, *Packeteable Implementation in Network Technlogy Series*. Cisco Press Networking. 2007

[7] Blum, R. C# Network Programming. John Wiley and Sons 2002.

[8] Cacti: http://www.cacti.net/

[9] Nagios: http://nagios.org/

[10] SmokePing: http:oss.oetiker.ch/smokeping/