Visualisation methods for real time and historical IP traffic monitoring [1]

Barry Irwin, George Wells

Department of Computer Science, Rhodes University
Grahamstown, 6139, South Africa

Abstract - **Various methods of graphical representation of IP traffic statistics over time are discussed. Methods are evaluated for applicability to near-real time monitoring and long term trend analyses are discussed. Implementations of various methods are also evaluated.**

1. Introduction

The ability for a network administrator to be able to quantify and qualitatively evaluate traffic on a managed network is an important need. Whether the network is a relatively slow WAN link or high speed LAN, being aware of traffic on the network is important for both troubleshooting and long term planning of the future development of the resource.

The network manager needs to be able to interpret the data collected from a variety of sources on a macro level in order to be able to spot trends in network usage, and for future capacity planning. Traffic data is usually gathered from routers or other intelligent networking equipment. In the raw numeric form in which it is gathered data is of limited use and needs to be processed into information before being interpreted.

Visualisation entails the transformation of information from a numerical tabular format to a graphical form. The graphical representation of the data allows for the concise display of large amounts of information. There is also the ability for certain information to be highlighted through the use of shape and colour.

Visual inspection of long-term graphical data allows for trends and anomalies to be seen with greater ease than if presented in tabular form.

Four classes of visualisation are discussed in section 3 along with some implementations of each. These same visualisation methods used for 'real' traffic can be used for predictive modelling and as an aid in answering 'what if?' scenarios. The underlying data can be modified to reflect the proposed scenario, and then remodelled.

2.1 Historic records

In order for longer term trend analysis and visualisation to be effective, data samples need to be stored over a period of time. The format in which data is stored should be such that it is easily accessible, both in terms of ease of use, and speed for searching. It has been found that use of a relational database, such as one of the many SQL-based engines, satisfies these requirements, as well as storing the data in a compact form.

The resolution and sampling frequency at which this data is stored impacts greatly on the quality of the visualisation and analysis possible, as well as the amount of storage space required. Hence a balance needs to be found between storage available, and the level of granularity of the sampled data. For most purposes sampling rates of one sample taken between every five and fifteen minutes should allow for reasonable processing of the data. Resolution can always be decreased during analysis, but it is impossible to increase the resolution of data after the original samples have been lost.

The keeping of long term data allows for after-the-fact processing, and the development of statistical models of traffic flow. Comparison of current data with historical data taken from the same period, or a statistically 'normalised' dataset, can prove valuable for planning.

2.2 Real-time and Near-Real time Monitoring

Traffic can be monitored in real time by a number of applications. These applications are useful for monitoring instantaneous traffic, but offer little in the way of longer term recording. Near real time applications, such as MRTG [6], display data current to the last sample interval (usually set to five minutes). This sample is also not a snapshot of the traffic at the time of capture, but rather the sum of traffic since the last sample was taken. Aggregation of traffic between sample intervals allows for some initial smoothing of data. Data captured for these near real-time

[1] This work was undertaken in the Distributed Multimedia CoE at Rhodes University, with financial support from Telkom, Lucent Technologies, Dimension Data, THRIP and the NRF

applications can also be logged to some form of long term storage, where they can be used as a base for later analysis. If traffic samples are too frequent and traffic is low, then very little can be gleaned from the resultant output. Near-real time visualisation and monitoring applications are sufficient for the majority of needs.

Real-time monitoring is best suited for specific troubleshooting or localised analysis. The extra load induced onto the network and network equipment by monitors gathering data from remote units, often impacts on the data collected.

3. Visualisation Methods

The following are four general classes of traffic visualisation that have been explored in research conducted so far. These are discussed in order of their complexity, and their applicability for certain types of information display is outlined.

3.1 Simple graphing

One of the simplest forms of visualisation, and one of the most commonly used is plotting a usage graph of the form used by Tobias Oetiker's MRTG and RRDTool packages [6,7]. Traffic samples are plotted as a percentage of total link capacity, derived from the result of averaging sample differences over time between samples, and alternately, the instantaneous throughput at the time of the sample. Incoming and outgoing traffic are usually plotted as two separate series.

The method used by these applications suffers from problems with representing data over long periods without substantial information loss as data is compressed and aggregated over time due to their design of having fixed size data files mostly for speed reasons.

Other implementations need not have such shortcomings. A system that provides a long-term data store using a database back-end can perform the same function at the expense of greater complexity and processing, yet it also retains flexibility with regards to the data to be displayed. Such a system would be able to provide the equivalent of the MRTG daily traffic usage, but from two months or even a year previously.

A web-based implementation along these lines using Perl, PHP [9] and MySQL [4] which allows a user to browse daily, weekly and monthly graphs for any period of time (dependant on data available in the database) has been constructed. The images are dynamically generated depending on the user's request. This form of visualisation is one of the most frequently used, and can provide an indication of when more in-depth analysis should be performed.

3.2 Layered graphs

While a simple graph is useful for monitoring spikes in traffic or link utilisation, it is not easily apparent as to what the cause of the increase in traffic is. One solution would be to have multiple graphs for various traffic sources and/or protocols, which can be further subdivided (as in the case of IP traffic to TCP or UDP ports) [3]. These graphs can then be compared to the overall graph, and patterns matched up. An alternative would be to plot these graphs representing fractional portions of total traffic as a cumulative graph, with the values stacked. The major use of such visualisation aids would be for obtaining more detailed and informative views of the traffic peaks as indicated by simple graphing formats, allowing an administrator to quickly see what is using the majority of the resource being monitored.

A major disadvantage of such a system is the complexity of assigning appropriate 'bins' for performing traffic accounting, since the number of 'bins' monitored is inversely proportional to the readability, and subsequent usefulness of output [1]. For example, the TCP and UDP components of IP each allow use of 65535 ports. Interpretation of a graph showing nearly 130 000 'bins' would be impossible, since even if coloured differently, the human eye would have trouble distinguishing the colours. Even if the majority of these ports had no traffic recorded, and hence not displayed, the output would still be confusing. Limiting the output to less than twenty variable 'bins' allows for better interpretation.

The advantage of use of this type of visualisation is that once 'bins' have been appropriately set up (usually on commonly used 'high bandwidth' ports such as FTP, HTTP, NNTP, SMTP and a catchall for non-specified traffic), in most it is cases fairly apparent which 'bin' is constituting the major portion of network traffic. The protocol, once identified, can then be managed using appropriate means. [2]. This format is also very useful for simulating the outcomes of adding various bandwidth reduction techniques such as caching proxy servers, as the composition of traffic after perceived savings can be clearly represented.

The current format of the application developed as part of research conducted, makes use of Perl as glue to gather data from a variety of sources (tcpdump, firewall accounting statistics, or database). Data is then processed, and rendered an image for viewing via a web interface, which also allows specific traffic

selection. The database is populated both from firewall accounting counters and from other Perl based counter applications collecting data from the raw network traffic using support programs such as tcpdump and RMON probes. Through the use of this approach graphs can be generated showing usage by protocol, as well as by IP address or group of IP addresses.

3.3 Colour Maps

This type of visualisation requires substantially more data than daily or instantaneous samples and is more suited for much longer term trend analysis. Colour map plots are generated as 2D projections of a 3D mesh representing traffic measurements over a particular period. The resultant plot is a height field with false colouring based on 'height' of the sample. High usage periods and other events occurring over a period of time can be distinguished due to the colouring. These are useful for much longer term historical analysis such as monthly or even annual summary. Areas of interest can be identified, and other tools used to view the details of traffic, provided that historical data has been kept in sufficient detail. Using such a method, trends can be identified over a longer time period, such as high traffic loads on a particular day of the week or month. These are more difficult to identify using traditional (x,y) type representation. Of particular interest is the organisation of data, aligned on day of week. This in most cases shows decreased traffic over weekend periods. Aligning on time of day is also useful for trend analysis at a lower level, but is particularly suited for 3D landscape analysis.

An alternative to the height-field projection method is to use a false colour projection of the contours of a 3D mesh. This is often more useful in determining particular recurring spikes in traffic. The rate of change is also implied through the use of contour mapping, with tightly spaced contours indicating rapid changes.

For testing purposed a utility was developed, using Perl to extract selected records from the traffic database, and then provide this reformatted data to GNUPlot [10] for rendering to either contour or colour maps. The resultant colour map image was displayed by means of a web-based interface to the system.

3.4 3D Landscapes

Using longer term historic data, 3D landscapes can be rendered which can be examined for various features such as valleys, plateaus and crevices. Such landscapes should be navigable, to allow for closer inspection. These essentially present the same data as colour maps, but with the added benefit that the administrator is able to clearly see the 3D nature of the data. The ability to 'navigate' round the landscape allows the viewer to see features that might otherwise be hidden. A further enhancement is to provide coloured strata to compose the landscape, with each colour representing certain bandwidth levels. Using coloured strata, enables the viewer to easily determine the level of bandwidth usage indicated by a particular feature, and in particular when there are valley's or troughs what the value at the bottom is.

The test system developed, utilises Perl for the extraction and formatting of traffic data, which is passed to the POVRay [8] rendering engine. The resultant landscape can be displayed as a single image, or a flyover can be selected to allow more detailed viewing of the landscape.

The disadvantage of this system is that the rendering is time and processor intensive, and as such is not particularly suited for real-time interactive viewing. Determining the exact date/time related to traffic illustrated using this method, can be difficult. This method is best suited for gaining an overall impression of traffic flows.

4. Conclusion

Historical records of data can be used for identifying important trends in network traffic. Visualisation models allow for administrators to get the 'big picture' of what is happening, and then use other tools to diagnose actual details. Simpler methods are more suitable for real time monitoring, where only fairly recent data is needed for a visual comparison. In order to display the larger amounts of data, associated with longer term analysis other visualisation methods are more suited.

Initial results, in particular colour maps and the landscape generation, have shown some considerable promise for trend analysis by visual means. Layer visualisation is more useful for after the fact diagnostics as to what was constituting heavy traffic loads as identified using the longer term and near real time methods. The layer graphing seems most valuable for identifying heavy traffic loads such as HTTP and NNTP, Napster and other file sharing protocols. Modelling the possible results of the introduction of caching proxies or filtering for such protocols can be rapidly visualised and qualitatively evaluated.

The level of success in using visualisation methods, for performing traffic analyses, depends largely on using the right form of visualisation to identify trends and anomalies. It is suggested that for long term analyses

that the more complex colour map and landscape methods be tried first, and the simpler 2D graphs used for specific investigation.

Tools for implementing methods described above are still under development and are undergoing refinement and optimisation. Only tools for the simple graphing category of visualisation have been found. Examples of this and the other classes of visualisation have been implemented and evaluated as part of the research carried out. An application could conceivably be constructed which operated on a single unified data store, and provided an interface for visualising this store through the methods provided, in addition to providing access to numerical data where required.

All of these methods can be used with varying degrees of success for near-realtime and long-term data visualisation.

References

[1] B. Irwin, "An analysis of common high bandwidth protocols", http://rucus.ru.ac.za/~bvi/research/papers/proto_analysis.txt, 2000.

[2] B. Irwin, "Reclaiming one's bandwidth: Dynamic filtering of traffic based on packet payload content", SACLA 2000 proceedings p105-109, http://rucus.ru.ac.za/~bvi/research/papers/sacla2000.pdf, 2000

[3] B. Irwin, "A per protocol approach to bandwidth monitoring and management of IP traffic" http://rucus.ru.ac.za/~bvi/research/papers/per-protocolbandwidth.ps, 2000.

[4] MySQL Open Source SQL database, http://www.mysql.org

[5] Napster.com, Napster Homepage, http://www.napster.com

[6] T. Oetiker, D. Rand, "MRTG: The Multi Router Traffic Grapher" http://www.mrtg.org

[7] T. Oetiker, "Round Robin Data Tool", http://rrdtool.eu.org

[8] Persistance of Vision Raytracer, http://www.povray.org

[9] PHP.net "PHP Hypertext Preprocessor" http://www.php.net

[10] T. Williams, C.Kelley, "An Interactive Plotting Program", http://www.gnuplot.org