Characterizing Network Performance for Enterprise Networks

Kurt Tutschku† and Herbert Baier‡

†University of Würzburg
Institute of Computer Science
Am Hubland, D-97074 Würzburg, Germany.
Tel.: +49-931-8886641, Fax: +49-931-8884601
e-mail: tutschku@informatik.uni-wuerzburg.de

‡InfoSim GmbH

Sedanstr. 27, D-97082 Würzburg, Germany.
Tel.: +49-931-4194590, Fax: +49-931-4194589
e-mail:baier@infosim.net

Abstract: This paper outlines first the need for a performance focused intranet management framework and second, introduces a new performance metric, denoted as "Network Comfort". Network Comfort characterizes the system performance by normalizing the instantaneous performance with the best one seen during network operation so far. Network Comfort maps the performance to an easy to recognize value in the interval from zero to one. In this way, the Network Comfort is capable to describe the elasticity present in IP networks and facilitates a transparent performance management. Furthermore, the concept is extended to application and service management.

I. Introduction

In recent years, the challenges to enterprise network management have grown tremendously. Mainly, three developments account for this:

a) the enterprises are implementing IP (Internet Protocol) based *intranets*. IP systems are easy to configure, simple to extend and integrate a wide variety of applications. IP networks, however, provide only limited configuration and performance control mechanisms. As a result, configuration changes occur more often and performance problem are difficult to identify.

b) information processing is becoming the core business of many companies. For examples, the business case of on-line brokers works only under the prerequisite of a high performing network. The ratio of employees working with these networks and number of handled transaction is amazingly high. As a result, the network management has to focus strongly on performance issues.

c) the debut of new enterprises structures. The companies form small business entities which are more flexible and permit a better cost allocation. This influences intranet management in double regard. The management itself is often out-sourced and therefore has to show its efficiency. In addition, network management has to cope with a high dynamic in the network configuration due to permanent company re-organization.

As a result of these developments, the approach to IP based intranet management has to be changed. Modern intranet management has to focus strongly on performance and service management under the constraint of operating an IP based system. Therefore a simple and appropriate performance characterization, addressing the specific features of IP systems, is indispensably for intranets. Such a characterization method will be presented in this paper.

The paper is organized as follows: in Section II, first, the need for a focused intranet management framework is discussed, and second, the new requirements on performance characterizations are stated. In Section III, the concept of *Network Comfort* will be introduced, which is a simple and efficient metric for characterizing network performance in intranets. Section IV investigates how the concept of Network Comfort can be applied to network applications. Finally, Section V will give a summary.

II. Intranet Performance Management

A. A Focused intranet Management Framework

Several network management frameworks like the *ISO/OSI management model*, cf. [1], or the *Telecommunication Management Network (TMN)* concept, cf. [2], have been developed in the past years. These architectures cover comprehensively a wide variety of management tasks. Particular the *FCAPS* taxonomy of functional areas, and consisting of *F*ault Management, *C*onfiguration Management, *A*ccounting Management, *P*erformance Management, and *S*ecurity Management, is well accepted and has been implemented in a number of network management tools, e.g. HP OpenView Network

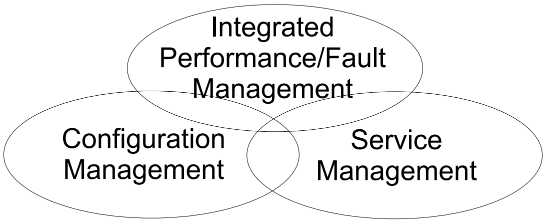

Fig. 1. Functional areas for performance-focused network management

Node Manager, cf. [3]. Although these frameworks are theoretical profound, the application of the concepts is still limited. Based a on careful analysis as well as from experience in projects, cf. Section III-E.1, three main reasons can be stated for this:

First, the frameworks are too complex. Often, only a subset of the functional areas is needed or applicable. Thus, network administrators prefer isolated approaches. In this way, minimizing their effort while increasing their efficiency.

Second, the interaction between the functional management areas of future high-performance networks are not appropriately addressed in the frameworks. The separation of the areas Performance Management and Fault Management tempt to address the common tasks of these functional areas in distinguished ways. For example, *soft errors*, indicating performance degradation due to faults, need an integrated approach for being identified and resolved.

Third, important business processes in network management are not adequately represented in the above mentioned frameworks. Especially, specifying, monitoring and reporting of *service levels* and *network quality* are strongly neglected. Proving the efficiency of network operation and network management is essential in intranets.

Thus, the need for a more focused network management concept arises. To address the challenges of operating high-performance IP based intranets, we propose a management framework focusing on three tasks: a) *Configuration Management*, b) *Integrated Performance/Fault Management*, and c) *Service Management*, see Figure 1.

It should be mentioned here, that the authors don't deny the importance of other management areas, particular Security Management or Accounting Management. In intranet management, however, these issue are often strongly decoupled, e.g. security is provided due to closed systems, or not applied, e.g. flat rates for connecting company divisions to the intranet.

Configuration Management

Configuration management constitutes the basis for efficient network operation. It provides methods to identify, adapt, and maintain element configuration, cf. [4].

In order to facilitate high performance and service-oriented network operation, configuration management has to extend its scope beyond the single element. Modern configuration management is required to support network-wide administration issues, like *traffic engineering* in future MPLS networks, cf. [5] or *network policies*, cf. [6]. Thus, Configuration Management methods have to be developed with regards to these issues. Network-wide configuration management is facilitated by the application of network object models like the *CIM (Common Information Model)*, cf. [7], and the concise mapping of the network objectives to the element parameters.

Since companies re-organize continuously their structure and therefore their communication network, a particular challenge in configuration management is keeping up with frequent element, topology and policy changes. Reaching this aim is partly facilitated using sophisticated network discovery mechanisms, cf. [8]. However, the discovery mechanisms still require some minimal configuration on the systems, e.g SNMP access should be feasible. Often, this is not possible due to incomplete element installation. The aim of having complete and up-to-date configuration information can be hardly achieved. Network management methods always have to account for incomplete knowledge.

Performance/Fault Management

Conventional Performance Management involves tasks like performance monitoring, problem isolation, performance tuning, analysis of statistical data for recognition of trends, and resource planning, cf. [4]. Traditional Fault Management comprises fault detection, fault location, service restoration, identification of the problems' root cause, and problem resolution. Since Performance Management aims at similar objectives as Fault Management, it can be view as the consequent extension of Fault Management, cf. [9].

We argue that, this view has to be changed in IP based networks. One reason for this is that these systems are already inherently fault tolerant due to dynamic routing protocols, e.g. RIP-2, cf. [10], or OSPF Version 2, cf. [11]. These legacy IP routing protocols do not guaranteed to preserve the performance characteristics completely, e.g. the provisioned bandwidth. This example outlines the complex interaction and multiple dependencies of mechanisms in IP networks. Thus, a strong interaction of Fault and Performance Management is indicated in order to resolve these problems during operation.

Therefore, we suggest that Fault Management should be viewed as a special case of Performance Management.

Service Management

The main task of *Service Management*, as defined here, is the integration of business processes related to network operation, e.g. establishing service level agreements, into the technical processes of engineering the network. Service Management comprises issues like service levels specifications and negotiation, monitoring the service quality and reporting the performance in an appropriate way. These functional tasks are partly addressed in conventional Performance Management, cf. [9]. However, more elaborated mechanisms and procedures are required for intranets, particular with multiple administrative domains. To tackle this task, models and protocols for automatic service level negotiation have been proposed recently, cf. [12]. Additionally, easy-to-understand methods for visualizing the network quality are needed. Such a methods is one of the main contributions of this paper and presented in detail in Section III.

SNMP network management architecture

The investigation of IP network management procedures is incomplete without the discussion of the *Simple Network Management Protocol (SNMP)*, cf. [13]. The SNMP concept defines an organization and information model and a communication protocol. However, there is no formal specification of the functional areas in SNMP network management.

The main advantage of SNMP is its widespread availability and interoperability between agents and managers of different vendors. The major weakness of the SNMP concept is the limited or missing specification of functional areas. SNMP provides mainly basic mechanisms for element management. Another weak point of SNMP is its "in-band" management concept: the control information is transmitted on the same channel as the user traffic. This implies that, *a)* fault and control information cannot be transmitted in certain error scenarios, and *b)* the management traffic puts always additional load on the network.

These inherent drawbacks of SNMP should be kept in mind, when proposing new management mechanisms for IP networks.

B. Characterizing Network Performance

B.1 Conventional Network-level Performance Metrics

Today's applied network-level performance metrics address two main performance aspects, network availability and network responsiveness, cf. [14]. The network availability can be characterized by metrics like the *network connectivity*, the *outage count*, the *error rate* and the *packet loss probability*.

Common network-level responsive metrics are *one-way delay*, *roundtrip latency*, *delay jitter* and *allowable bandwidth*.

The above enumerated performance metrics, particular the delay, delay jitter and packet loss probability, underline strongly the technical character of the metrics. They have been designed mainly to describe the efficiency of the basic packet transport.

B.2 Additional Requirements on Network-level Performance Characterization

In the context of the focused management framework outlined in Section II-A and the so far applied network-level performance metrics, additional requirements on characterizing network performance can be stated now.

These additional requirements emerge mainly from the characteristics of IP networks and from the increasing interaction of network users and networks operators.

Network performance characterization should be *unambiguous* and *simple* in order to facilitate fast Performance/Fault management. Hence, technical metrics must be naturally the basis of the characterization. In addition, the characterization should facilitate the *identification of operational modes*, for example the traffic light notion: green for good condition, yellow for attention needed and red for critical state.

In relation to the above stated requirements, network operation demands the description of the *persistent behaviour* of the system. Therefore, performance metrics should be enhanced by a *temporal component*.

Performance characterization should be *easy-to-understand*. Technical values should not dominate the description. In this way, enabling non-IT people the understanding of the system's behavior and in consequence, making Service Management more transparent.

Performance indicators should reflect the *user's perception* of the system in an objective way. It should characterize the quality of the consumed service in terms of the user. Thus, providing less ambiguity and better support for Performance and Fault management. However, due to the unknown baseline used for comparison, characterizing a subjective quality value in an objective way is hard to achieve. A favorable solution to this problem is the approach to *normalize* the momentary seen quality by the best-ever recorded performance. The normalization concept is one of the main contributions of the work presented herein, cf. III.

In addition, network performance characterization should directly describe the *system degradation* due to high utilization, e.g. it should discriminate the additional amount of packet delay generated by longer queuing times. Thus, the characterization permits the network administrator to locate points of congestion.

Furthermore, the performance characterization should describe the *amount of elasticity* present in IP networks. *Elasticity* is a key performance characteristic in these systems, resulting from the application of the *Transmission Control Protocol (TCP)*, over which most of the data is transported, cf. [15].

The second set of requirements on performance metrics origins from their implementation needs. Due to the in-band transmission of control information in IP based networks, the monitoring of an metric should impose *minimal additional load* on the network. Additionally, monitoring should require *no or only less instrumentation* on the surveyed network elements. In order to facilitate wide-spread implementation, the monitoring should apply *vendor neutral* mechanisms and protocols. Of course, the monitoring should be *scalable*.

To facilitate the above stated additional requirements, a new performance characterization concept, denoted as *Network Comfort*, is proposed and will be discussed next.

III. Network Comfort

Network Comfort is derived from the conventional network-level metric *roundtrip latency*. It is based on round trip time measurements, which are normalized by the minimal recorded time of a certain type of packets. The normalized values are averaged over a time interval.

A. Definition

This leads to a two step definition. First, we define the *single network comfort measurement* κ_i at time t_i as:

$$\kappa_i = \frac{rtt_{\min,i}}{rtt_{\text{inst},i}}, \tag{1}$$

where the variable $rtt_{\text{inst},i}$ is the instantaneous measured roundtrip latency at time t_i from the measurement system to the network element under test and $rtt_{\min,i}$ denotes the minimal roundtrip time. The temporal scope of the minimum is of great interest and discussed in detail in Section III-C.

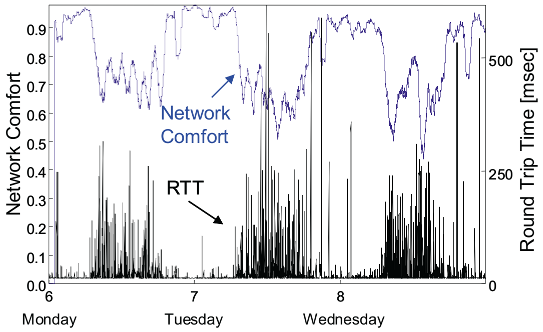

Fig. 2. Network Comfort of Router RBTRZ04

Now, the *Network Comfort* nc_i at time t_i is defined as:

$$nc_i = \frac{1}{W} \sum_{j=0}^{W-1} \kappa_{i-j}, \tag{2}$$

where W specifies the window size, i.e. the number of single measurements incorporated into the time average. The way of sampling the measurements, e.g. periodic or burst sampling, and the sampling interval are by purpose not specified in Eq. (2). The selection of this parameters is left up to the user. For a detailed discussion on periodic sampling and selecting the sampling interval see Section III-D.2.

An example of the Network Comfort is depicted in Figure 2. It shows measurements taken over a period of three days and is part of the case study presented in Section III-E.1. In this study ICMP echo request packets of size 64 bytes have been used, cf. Section III-B. In Figure 2, the black curve depicts the roundtrip time (RTT) from the measurement equipment to the monitored router RBTRZ04. As expected, the roundtrip time (RTT) fluctuates strongly and the daily variation within the business hours can be identified. The blue line depicts the corresponding *Network Comfort*. The behaviour of Network Comfort is much smoother and the states of increased network response times are more clearly visible.

B. Characteristics of Network Comfort

The term *Network Comfort* was selected with purpose. Besides underlining that it is a network-level metric, the term *comfort* emphasizes that the indicator describes the decreased convenience of using the path to the system under test.

The normalization of the instantaneous measured roundtrip time by the minimum latency yields a mapping of κ_i to the interval of $(0;1]$. A value of one corresponds to an optimal network performance. A value near to zero indicates an almost completely degraded network. In this way, Network Comfort facilitates the aim of obtaining an easy-to-understand description of path's performance degradation. The concept of Network Comfort is similar to the *Fun Factor* introduced in [16]. Whereas the Fun Factor is suggested for network planning, the Network Comfort is directly suited for network operation and performance management.

A key feature of Network Comfort is its capability to describe the elasticity present in the network. The normalization together with the computation of the time average permits the approximation of the extension of transmission times. A Network Comfort value of 0.5, for examples, indicates a doubled download time, a value of 1/3 a tripled download time.

The computation of the of moving average leverages short overload periods and characterizes *persistent* system states. Thus, the network administrator is able to identify and locate performance problems *unambiguously* and *reliably*.

Since Network Comfort is based on response time measurements, it reflects the user's perception of the networks quality. Furthermore, the normalization to the interval of $]0;1]$ is easy to read and it prevents from overweighing absolute technical values. In this way, Network Comfort facilitates the requirement that a performance metrics should be interpretable for non-IT people.

Another important characteristic of Network Comfort is its capability to describe the additional amount of delay introduced by high utilization. This comes clear from rewriting Eq. (1):

$$\kappa_i = \frac{rtt_{\text{min},i}}{(rtt_{\text{inst},i} - rtt_{\text{min},i}) + rtt_{\text{min},i}} = \left(\frac{1}{rtt_{\text{min},i}}(rtt_{\text{inst},i} - rtt_{\text{min},i}) + 1\right)^{-1} . \quad (3)$$

The difference $(rtt_{\text{inst},i} - rtt_{\text{min},i})$ denotes the additional delay generated by longer queuing times, seen by a test packet at time t_i.

Implementation characteristics

Network Comfort can easily be implemented in IP based networks by exploiting the *ICMP echo request/reply* mechanism, cf. [17]. Any IP machine receiving an echo request is supposed to respond with an echo reply. Thus, no additional instrumentation on IP network elements is required and measurements can be deployed rapidly. Furthermore, by using the ICMP protocol, the measurement concept is highly vendor neutral as demand in Section II-B.2.

Furthermore, using ICMP packets imposes only a low amount of traffic on the network. An echo request packet consists of 20 byte IP header and an eight byte ICMP specific part, which is followed by an arbitrary amount of padding data. Typically, ICMP packet of 64 byte size where used throughout experiments present in this work.

In addition, ICMP echo requests are already widely used in IP network management for topology discovery and active liveness test. The Network Comfort measurements can be piggy backed easily on this procedures.

C. Scope of Minimum Determination

The minimal roundtrip time value rtt_{min} can be determined in the context of two temporal scopes, an *off-line* mode and an *on-line* mode.

In the *off-line mode*, perfect knowledge over all measurements is assumed. Hence, $rtt_{\text{min},i}$ at time t_i is defined as:

$$rtt_{\text{min},i} = \min_{\forall}\{rtt_{\text{inst},k}\}; \quad (4)$$

note that Eq. (4) does not depend on i. The off-line mode is suggested when the measurements are performed remotely and their analysis is conducted separately. In the off-line mode, the performance characterization of the Network Comfort is always correct due to the knowledge of the global minimum. In real network operation, however, a performance characterization should be performed immediately, and the application of this mode is limited.

In the *on-line mode* the minimum round trip time is learned during network operation. Here, $rtt_{\text{min},i}$ at time t_i is defined as:

$$rtt_{\text{min},i} = \begin{cases} rtt_{\text{inst},1} & \text{if } i = 1 \\ rtt_{\text{inst},i} & \text{if } (i > 1) \wedge \\ & rtt_{\text{inst},i} < rtt_{\text{min},i-1} \\ rtt_{\text{min},i-1} & \text{otherwise} \end{cases} \quad (5)$$

Eq. (5) specifies, in contrast to the off-line mode, a valid minimum for every time instance during the monitoring process. In this way, permitting the application in daily operation.

Learning the minimum over time, however, introduces a temporal ambiguity in the Network Comfort measure. Network Comfort values of different minima should not be compared. Thus, the on-line mode is sophisticated but requires careful interpretation.

Two sources of temporal ambiguity can be identified while using the on-line mode. The first source is resulting from configuration, topology and routes changes. The second source is the instationarity inherent to the learning process.

Impact of configuration, topology and route changes

Configuration, topology and route changes have to be incorporated immediately in order to provide a valid characterization of the network's performance. This is assured by Eq. (5), since at every measurement, the minimum value is compared with the instantaneous round time and updated if necessary.

The impact of route changes, e.g. due to load balancing, fault recovery operations or network optimization, should be reflected in the characterization in such a way that the best route can be identified. Eq. (5) guarantees this, since the global minimum is kept, once it is obtained. In this way, route changes which lead to an increase in the round trip time and, in turn decreasing the Network Comfort, can be identified easily.

Impact of the observation process

The second type of ambiguity while using the on-line mode is the instationarity inherent to the learning process. Obtaining the minimum requires some amount of time. During this period the Network Comfort values should not be compared with each other. To proof the applicability of the on-line mode, it is necessary to investigate the convergence behaviour of the minimum under assumption that no configuration change has occurred. In particular the question how fast the global minimum is obtained has to be answered.

Figures 3 shows the convergence behaviour of the minimum round trip time for router RBTRZ04, cf. Section III-E.1. No configuration change occured during the observation period. The background part of Figure 3 depicts the long term behaviour of the minimum over a period of 21 days. On the first day, the so far observed minimum value drops significantly. This value is kept until day 5, when global minimum is obtained. The drop to the global minimum is only marginally and within a few tens of a millisecond. After day 5, the minimum stays on this level until the end of the measurement. The zoom-in part Figure 3 shows the behaviour of the minimum round trip time on the first day of the measurement. The value drops rapidly within the first six hours.

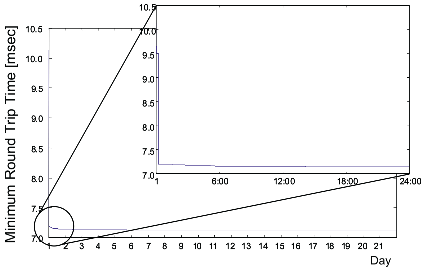

Fig. 3. Temporal Behaviour of the Minimum Round Trip Time to Router RBTRZ04

This empirical analysis indicates that the online mode is applicable in real network operation. The minimum is usually obtained within a few hours of measurement, a time for which ambiguity can be accepted.

D. Determination of Measurement Parameters

Measuring the Network Comfort requires from the network administrator only the determination of two types of the parameters, the selection of the window size and the specification of the sampling parameters.

D.1 Selection of Window Size

The size of the window for the moving average specifies the *sensitivity* of the Network Comfort measure on degradation phases. It should be selected such that the window reflects the expected *long term time constant* for the system.

From the user's perception, the *long term time constant* is the session length. From network administrator point of view, the *long term time constant* should be in the order of appropriate management actions. That means, the selection of the window size should support the network administrator, such that short overload periods are neglected while persistent degradation can be identified.

Determining the user's session length is a demanding task. A lot of research has already been carried out for public networks, cf. [18], or access networks, cf. [19]. However, only few efforts has been laid on IP based enterprise net-

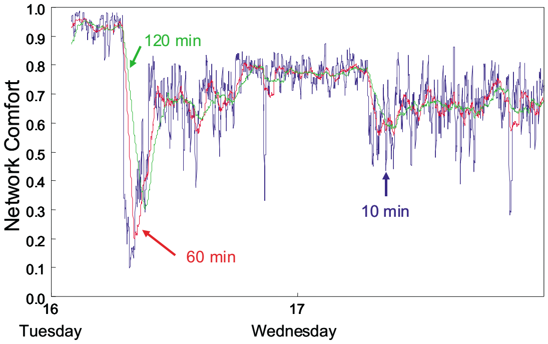

Fig. 4. Network Comfort Behaviour for a 10min, 60min, and 120min window

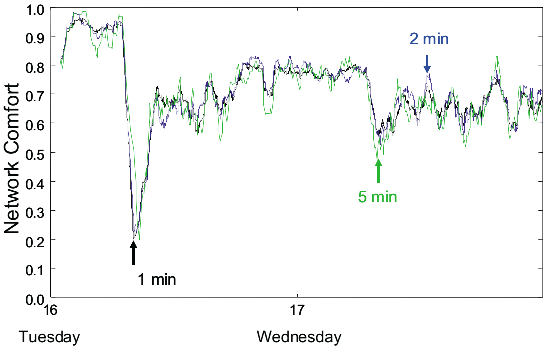

Fig. 5. Influence of different sampling intervals

works. Thus, for this study common results from public networks were extrapolated. It was assumed that the session length in an intranet is $1.5 - 2$ times longer than in public networks. Thus, resulting in a session duration in the order of 1hour. Starting from this assumption, three window sizes for the moving average have been investigated: 10min, 60min and 120min.

Figure 4 compares the Network Comfort to router RBTRZ04 for the three window sizes. The sampling interval between single network comfort measurements was 1min and regular periodic sampling has been applied. The blue line denotes the Network Comfort for the 10min average, the red line for a window size of 60min, and the green curve a window of 120min. The 10min average fluctuates strongly and permits hardly the identification of persistent phases. The 60min average is much smoother and shows clearly states of reduced network comfort. The 120min average is leverages very strongly. Figure 4 shows that a window size of 10min is too short and can not appropriately reflect the expected time constant. The 120min average reduces the sensitivity too much. In addition, the time lag of the 120min average becomes visible. A window size of 60min, however, seems to be well suited and reflect the time constant appropriately.

D.2 Choice of Sampling Parameters

The choice of the sampling parameters determines the *accuracy* of the Network Comfort measure. Two sampling parameter have to be specified, the sampling method and the sampling interval. A major constraint on selecting these parameters is the requirement that monitoring the Network Comfort should impose minimal additional traffic on the network.

Since Network Comfort characterizes the temporal behaviour of the path to the tested system, *periodic sampling* is suggested in order to reconstruct behaviour appropriately. Thus, the sampling interval is the main parameter to be defined.

In general, the sampling interval should be taken fairly shortly compared with the *long term time constant* expected for the system. Three sampling intervals of 1min, 2min, and 5min have been investigated for 60min time window . Figure 5 depicts the influence of different sampling intervals. The behaviour of the Network Comfort for the 1min and 2min intervals are very similar. The curve for the 5min sampling interval, however, shows a significantly stronger fluctuations. In this case, the sampling rate is too long and averaging over the time window can't leverage the extreme values appropriately.

At this point it is important to mention that beside the above introduced *long term time constant* also a *short term time constant* conducts the system behaviour. For a single network comfort measurement, cf. Eq. (1), the time constant is defined mainly by the waiting time of the packets in the queues along the path. Direct observation of the short term time constant by sending test packets in order of the round trip time is not feasible, since this would impose too much additional traffic. An realistic approach is probing the response time using *burst sampling*.

Burst sampling differs from periodic sampling such that at the sampling instant, a fixed number $N, N > 1$ of test packets, which form the *burst*, are issued back-to-back by a monitoring station. The individual round trip times of the test packets in the burst are recorded and averaged into a single measurement.

Figure 6 compares the Network Comfort to router BTRZ04 obtained by burst sampling with the one obtained for conventional periodic sampling. A 1min sampling interval and a burst size of five packets were used. Both curves show a similar behaviour and only small deviations are visible. The main difference between both curves occurs during phases of reduced Network

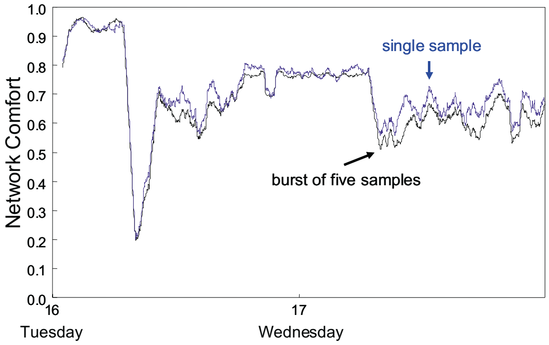

Fig. 6. Impact of burst sampling

Comfort. The reason for this is, that within a burst it is more likely that packets with larger round trip time are observed.

This comparison indicates that using burst sampling increases the accuracy of monitoring the Network Comfort. However, the application of conventional periodic sampling in daily network operation is still suggested. The small increase in accuracy doesn't justified that the significant rise of the additional network load, which is now five times higher.

E. Network Comfort Estimation

Because of its capability to describe elasticity and the amount delay in the network, the Network Comfort metric is interesting for network planning. For this purpose, the Network Behaviour has to be described analytically using the anticipated traffic as input.

In addition, Network Comfort is based on the observation of test packets. In real network environments, it is not guaranteed that the test traffic is handled in same way as the user traffic. Thus, the need arises to verify the Network Comfort measurements with analytic results.

To address these issues, a simple analytical model for the approximation of Network Comfort will be developed in Section III-E.2. The estimation will be compared with measurements from a case study of large intranet.

E.1 Case Study

Network Comfort is being used with success in a typical IP based intranet of a German insurance company, which has about 12,000 users distributed into approximately 300 locations.

Network Topology

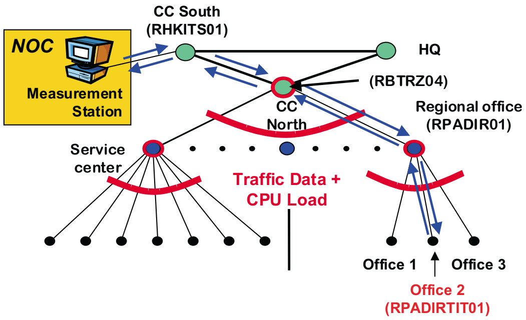

Fig. 7. Insurance company topology and the case study path.

The topology of the intranet is depicted in Figure 7. One Head-Quarter and two computing centers, named North and South, form a three corner backbone, whose links are 2Mbits leased lines. The network topology of the remainder locations is a tree of depth two, where:

a) the root is the computing center North;
b) the nodes with depth one are regional offices or service centers;
c) the nodes with depth two are local offices or hospitals.

The links between the computing center North and the regional offices or service centers are typically 2Mbits leased lines and between the regional offices or service centers and the offices or hospitals are either 64Kbits or 128Kbits ISDN leased lines. All in all, the network comprises approximately 600 router and switches.

Data Collection

The *network operating center (NOC)* is located in the computing center South, where a *measurement station* measures continuously round trip time data to various nodes spread around this intranet.

The traffic and load data for the routers are monitored by the measurement station continuously and with full network coverage. The traffic data are obtained by collecting periodically the MIB variables `ifInOctets` and `ifOutOctets`, denoting the total number of octets received and transmitted on an interface, cf. [13]. The CPU load on the routers was monitoring every 5min using the private CISCO variable `avgBusy5`, cf. [20].

E.2 Mean Value Approximation

The Network Comfort can be estimated using a mean value approximation for the single network comfort measurement. For the estimation it is assumed that the load on the routers can be neglected and that only the waiting time in the link buffers accounts for the delay. The waiting time in the buffers is modeled by a $M/M/1$ delay system.

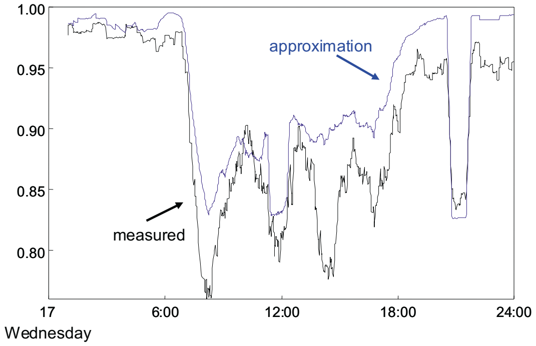

Fig. 8. Comparison of measured and approximated Network Comfort to router RPADIRTIT01

Thus, in Eq. (3) the term for the waiting time can be replaced by an approximation term:

$$\kappa_i \;=\; \left(\frac{1}{rtt_{\mathrm{min},i}}\left(W_{\mathrm{ins},i}\right)+1\right)^{-1}. \qquad (6)$$

with

$$W_{\mathrm{ins},i} \;=\; \sum_{\forall \text{traversed links } (j)} w_{\mathrm{ins},i}. \qquad (7)$$

The mean waiting time for a particular link is:

$$w_{\mathrm{ins},i}^{(j)} = (\rho_i^{(j)}/\mu^{(j)})/(1-\rho_i^{(j)}), \qquad (8)$$

with $\lambda_i^{(j)}$ as the arrival rate for link (j) at time instant i and $\mu^{(j)}$ as the service rate; and $\rho_i = \lambda_i^{(j)}/\mu^{(j)}$.

For the estimation, the arrival rate $\lambda_i^{(j)}$ is computed from the traffic data of the MIB variables observed at time instant i. The service rate $\mu^{(j)}$ is obtained from the known bandwidth of the link. The approximation of the single network comfort value is used for obtaining the time average of Eq.(2).

To evaluate the mean value approximation, the Network Comfort to router RPADIRTIT01 was investigated. A test packet traverses through the routers RHKITS01, RBTRZ04, and RPADIR01 on its path to RPADIRTIT01, cf. Figure 7.

Figure 8 compares the measured Network Comfort to router RPADIRTIT01 with the estimation. The observation period was one day. The black line shows the measured Network Comfort and the blue curve represents the approximation. The approximation follows the trend of measurements amazingly close. Only during phases of high performance degradation, the estimation deviates from the measurements.

This verifies that in the case study the test packets are handled in the same way as the regular traffic. Furthermore, the empirical evaluation shows that the simple mean value approximation is not numerical exact but provides a good qualitative anticipation of the Network Comfort trend. The good approximation indicates that the estimation can be used for coarse network planning tasks, often occurring in daily network operation.

In addition, in real network environments, full coverage data collection is often not possible. The approximation suggests that monitoring the Network Comfort is a simple but feasible alternative for network managers to evaluate performance characteristics.

IV. Web Comfort

An appealing characteristic of Network Comfort is its capability to describe elasticity on network level and to provide an objective measure for the user's perception. In this section, it is investigated whether this feature can be extend to network applications such as web hosting. Therefore, the notion of *Web Comfort* is introduced.

Similar to Network Comfort, the Web Comfort metric normalizes the instantaneous download time for a web page by minimal observed time. For the case of Web Comfort, the Eq. (1) is redefined as:

$$\omega_i = \frac{dlt_{\mathrm{min},i}}{dlt_{\mathrm{inst},i}}, \qquad (9)$$

where $dlt_{\mathrm{min},i}$ denotes the minimal *download time (dlt)* of the web page and the variable $dlt_{\mathrm{inst},i}$ is the instantaneous measured loading time. The single Web Comfort measurements are as well averaged for a time window, cf. Eq. (2).

The applicability of Web Comfort was tested by monitoring the download time of the web page "www.infosim.net/ag/index.html" for a period of two days. The page was downloaded very minute from a measurement station, using the public available "wget" utility, cf. [21]. The measurement station was located at University of Würzburg, Germany, and used no cache. The web server was at the InfoSim office, also located in Würzburg, Germany. A 60min window was used for averaging.

Figure 9 shows the observed download times, black curve, and the computed web comfort, blue line. The loading times fluctuate strongly but clearly show a daily trend. The Web Comfort provides a good identification of states of high loading time. Surprisingly, the trend is moved away from the business hours into the evening. This behaviour is explained by the fact that the data is routed over the Atlantic and back to Germany.

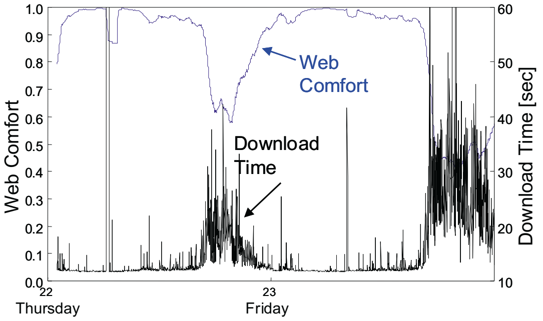

Fig. 9. Web Comfort of "www.infosim.net/ag/index.html"

The experiment indicates that the concept of Network Comfort can be extend to network applications revealing elasticity. The origin of elasticity, however, is often not unambiguous. It is either be build into the application, by using a client/server architecture, and/or comes from the use of the TCP transport mechanism. Thus, active measurements and synthetic transactions together with concepts like the Web/Network Comfort will become important to characterize the user perceived performance.

V. Conclusion

This paper outlined first the need for a performance focused intranet management framework and introduced second, a new performance metric, denoted as "Network Comfort". Network Comfort characterizes the system performance by normalizing the instantaneous network performance with the best one seen during network operation so far. Network Comfort maps the system performance to an easy to recognize value in the interval from zero to one. In this way, the Network Comfort is capable to describe the elasticity present in IP networks and facilitates a transparent performance management.

Furthermore, the concept was extended to application and service management. As an example, the notion of "Web Comfort" was introduced, which describes the download performance of a web page.

Future research has to be directed to a more elaborate analysis of the temporal characteristics of Network Comfort. Particular, towards the specification of the window size and the sampling parameters.

References

[1] ISO7498-4, "Information Processing Systems – Open Systems Interconnection – Basic Reference Model – Part 4: Management Framework," Technical reference, International Organization for Standardization, 1989.

[2] D. J. Sidor, "TMN standards: Satisfying today's needs while preparing for tomorrow," *IEEE Communications Magazine*, vol. 36, no. 3, pp. 54–64, March 1998.

[3] J. Huntington-Lee, K. Terplan, and J. Gibson, *HP OpenView – A Managers's Guide*, McGraw-Hill, New York, NY, 1996.

[4] M. Subramanian, *Network Management – Principles and Practice*, Addison-Wesley, Reading, Ma., 2000.

[5] D. Awduche, J. Malcolm, J. Agogbua, M. O'Dell, and J. McManus, "Requirements for traffic engineering over MPLS," Request for Comments 2702, Internet Engineering Task Force (IETF), 1999.

[6] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, J. Perry, S. Herzog, A.-N. Huynh, M. Carlson, and S. Waldbusser, "Policy terminology," <draft-ietf-policy-terminology-01.txt>, Internet Engineering Task Force (IETF), 2000.

[7] DMTF, "Common Information Model (CIM)–Specification Version 2.2," Technical reference, Distributed Management Task Force Inc., 1999.

[8] Y. Breitbart, M. Garofalakis, C. Cliff Martin, R. Rastogi, S. Seshadri, and A. Silberschatz, "Topology discovery in heterogeneous ip networks," in *Proceedings of the IEEE Infocom 2000*, Tel-Aviv. Israel, 2000, IEEE.

[9] H.-G. Hegering and S. Abeck, *Integrated Network and System Management*, Addison-Wesley, Reading, Ma., 1995.

[10] G. Malkin, "RIP Version 2," Request for Comments 2453, Internet Engineering Task Force (IETF), 1998.

[11] J. Moy, "OSPF Version 2," Request for Comments 2328, Internet Engineering Task Force (IETF), 1998.

[12] Y. T'Joens, D. Goderis, R. Rajan, S. Salsano, C. Jacquenet, G. Memenios, G. Pavlou, R. Egan, D. Griffin, P. Vanheuven, P. Georgatsos, and L. Georgiadis, "Service level specification and usage framework," <draft-manyfolks-sls-framework-00.txt>, Internet Engineering Task Force (IETF), 2000.

[13] W. Stallings, *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2*, Addison-Wesley, Reading, Ma., 1999.

[14] D. Verma, *Supporting Service Level Agreements on IP Networks*, MacMillan Technical Publishers, Indianapolis, IN, 1999.

[15] J. Charzinski, "Fun factor characterization of user perceived quality of service for elastic internet traffic," in *Proceedings of the KIVS 2001*, Hamburg. Germany, 2001, IEEE.

[16] J. Charzinski, "Fun factor dimensioning for elastic traffic," in *Proceedings of the ITC Specialist Seminar on Internet Traffic Measurement, Modeling and Management*, Monterey. CA., 2000, IEEE.

[17] J. Postel, "Internet control message protocol," Request for Comments 762, Internet Engineering Task Force (IETF), 1981.

[18] J. Kilpi, "Call level traffic analysis of a large ISP," in *Proceedings of the ITC Specialist Seminar on Internet Traffic Measurement, Modeling and Management*, Monterey. CA., 2000, IEEE.

[19] N. Vicari and St. Koehler, "Measuring internet user traffic behavior depenednt on access speed," in *Proceedings of the ITC Specialist Seminar on Internet Traffic Measurement, Modeling and Management*, Monterey. CA., 2000, IEEE.

[20] P. L. Della Maggiora, Ch. E. Elliott, R. L. Pavone, K. J. Phelps, and J. M. Thompson, *Performance and Fault Management*, Cisco Press, Indianapolis, IN, 2000.

[21] H. Niksic, "GNU Wget - the noninteractive downloading utility," available at http://www.gnu.org/software/wget/wget.html, Free Software Foundation, 1998.