A Method for Upper Bounding Long Term Growth of Network Access Speed

Thomas Phillip Knudsen*, Ahmed Patel, Jens Myrup Pedersen*, Ole Brun Madsen***
***Department of Control Engineering, Aalborg University**
Aalborg East, DK 9220, Denmark
****CNDSRG, Department of Computer Science, University College Dublin**
Belfield, Dublin 4, Ireland

ABSTRACT

The development in home Internet access speed has shown an exponential development with growth rates averaging 25% per year. For resource management in network provisioning it becomes an urgent question how long such growth can continue. This paper presents a method for calculating an upper bound to visual content driven growth, proceeding from datarate requirements for a full virtual environment. Scenarios and approaches for reducing datarate requirements are considered and discussed. The presented figures for an upper bound on network access speed are discussed and perspectives on further research presented.

Keywords: Computer Networks, Network Access Speed, Network Planning, Network Provisioning, Large Scale Networks.

INTRODUCTION

The development of Internet over the past years shows a marked growth in traffic. This development mirrors a comparative growth of speed in the typical network access connection, and datarates are swiftly approaching the absolute limit for capacity in the old twisted pair wire access networks. One of the most demanding application types is live image transfer, and the MPEG2 standard gives specifications for usage profiles ranging up to 100 megabit/s [1], well above the capacity for the Xdsl technologies. The growth rates for Internet connection access speed has been around 25% per year. Jakob Nielsen has formulated this growth tendency as Nielsens Law for NT access speed [2] predicting continued growth of 25% per year. Extending this growth in access speed are shown in Table (1).

Table 1 Projected network access speed

Years from now:	Speed:
Now	1 Megabit/s
10	10 Megabit/s
20	87 Megabit/s*
35	2.5 Gigabit/s
50	70 Gigabit/s
75	18 Terabit/s
100	4.9 Petabit/s

* suffices for MPEG2 high level-main profile.

In 1032 years there will be 1 Megabit/s for each of the estimated 10^100 nuclear particles in the universe!

Higher growth rates figure in the strategic plan published by the Swedish IT commission recommending a development from 5 megabit/s in 2005 with annual doubling up to 100 megabit/s [3]. This development will necessitate construction of a new access network structure based on optical fiber.

This growth must be bounded at some point, but it is by no means to be assumed that this will happen at 100 megabit/s. This raises the question when this exponential growth rate will cease; a question of major importance for resource management in dimensioning and planning for the future FTTH (Fiber To The Home) access nets and actual upgrading of the MAN-WAN. This is particularly pertinent as dimensioning and structuring of such networks determine the capacity for several decades. Subsequent upgrades involving laying down further tubing and cabling will have a cost comparable to the construction of the networks in the first place. An estimate of the upper bound for network access speed can be reached by determining the limit beyond which further growth would be absurd. This limit is here derived from analysis of the human capacity for utilizing bandwidth, where the bandwidth needed to provide a full virtual experience in real time constitutes the upper bound; beyond this limit no further increase in quality will be perceived by an observer, and content-driven growth will end.

This paper presents a method for calculating the upper bound on growth of network access speed based on technical requirements for supporting a full virtual environment to a human observer. The analysis takes as its premises the global Internet and a scenario of providing full real time telepresence with viewing of arbitrary content. The analysis deals with a single human observer. The analysis goes through three main steps, where first the premises are given, the global Internet and the human visual capabilities, followed by calculation of the gross bound from these premises, then different methods for reducing such bandwidth are introduced, and finally estimates of the bound are given and discussed.

PREMISES

To establish an upper bound based on premises, these must remain valid for the considered period. Most factors involving network content are volatile and unpredictable; The human eyesight on the other hand is a constant factor and can be characterized by a well defined set of factors - resolution, motion perception and colour depth. Taking this as the premise, specifics of content can be dispensed with, and an upper bound can be established on the basis of the technical requirements for providing a full real time visual environment for a human observer. Once the image quality has reached the limit for what is perceivable, any further increase in quality will be redundant. This quality limit is defined as visual reproduction in real time of a different arbitrary locale, with quality indistinguishable from direct, on site viewing. The visual reproduction is characterised by the data that must be presented to the observer; there are no assumptions made concerning the screen technology used to present it, except that it is not goggles worn by the observer; developments in screen technologies are expected to mitigating it as a limiting factor [5], and no assumptions about screens are necessary. Based on these premises, two basic scenarios are to be considered, a static observer and a moving observer:

Static Observer
By a static observer is meant an observer who is static in relation to the screen surfaces displaying the live image. As an example of full visual reproduction can be taken a room covered in screen surfaces displaying a live sports event, giving the observer the full experience of being present at the event. The detail level in the transmitted image is not a function of the screen surfaces, but of the acuity of the observers eyesight. Given the characteristics of human eyesight, the necessary detail level of the virtual environment can be calculated. If the observer is moving in relation to the screen surfaces, such as approaching, the image should zoom in and increase detail level on the approached screen surface retaining perceived image quality and adjusting for viewing angles.

Moving Observer
This moving scenario has the added complexity that with changing viewer location the image presented must be adjusted; thus the transmission must compensate for transmission loop latency. To the extent that observer movement can not be predicted it is necessary to transmit extra data - buffering in the image - enabling adjustment to be computed at the observers location. For upper bounding global transmission delays must be taken into account. These scenarios give the gross value for access speed; to reach the net figure known methods for compression must be taken into account. Such methods can be separated into three categories: non-destructive compression, destructive compression and Retinally Reconstructed Images (RRI) [6], where information is removed from the images based on the varying ability of the human eye to discern details.

Human Vision
The characteristics of normal human vision have long been established and are given here as normal eyesight. It is characterised by having a high angular resolution near focus and lower resolution as the angle from line of sight increases. In addition to this, newer research has demonstrated heightened visual sensitivity to certain phenomena such as swiftly blinking lights; this sensitivity, termed hyperacuity, is observed in both increased angular resolution and time resolution [9][10][11]. These characteristics are given in Table (2).

Table 2 Characteristics of human eyesight [6][9][10][11]

	Normal sight	Hyperacuity
Max angular resolution	1´	10´´
Time resolution	50 Hz	500 Hz
Colour depth	24 bit	24 bit

*involving changes in colour and space.

To display a given resolution image point separation must be half that.

SIMPLE MODEL

Static Observer Model
The simplest model is a sphere surrounding the observer having a granularity equivalent to the maximum angular resolution of human eyesight. To present a full virtual experience an image with this resolution must be transmitted. Two sets of figures are given, one a full sphere allowing arbitrary direction of focus without tracking, and a 160° by 90° section approximately corresponding to the visual area with constant direction of focus. In both cases each granular point contains 24 bit colour information, see Table (3) and Table (4).

Table 3 Number of points on sphere

Angular resolution:	Standard ½'	Hyperacc. 5"
Full Sphere	5,9e+8	5,9e+10
160° by 90°	2.1e+8	2.1e+10

Table 4 Bits/s at 24 bpp colour and 50 Hz update

Angular resolution:	Standard ½'	Hyperacc. 5"
Full Sphere	7,1e+11	7,1e+13
160° by 90°	2.5e+11	2.5e+13

Taking the 500Hz temporal hyperacuity into account the figures in Table (4) must be multiplied by ten.

Moving Observer Model
If the future position of the observer can not be predicted, a further level of complexity is added to the model. In this case sufficient information must be added to the transmission to enable retention of full image quality for all possible positions of the observer during the time period until the transmission can be adjusted. Thus, the access speed limit becomes dependent on the delay in the transmission control loop. This is illustrated by Fig. (1), where the inner circle represents the present observer position and the outer circle represents the limit of movement before transmission can be updated. The image must contain sufficient extra details to maintain resolution at all possible observer positions.

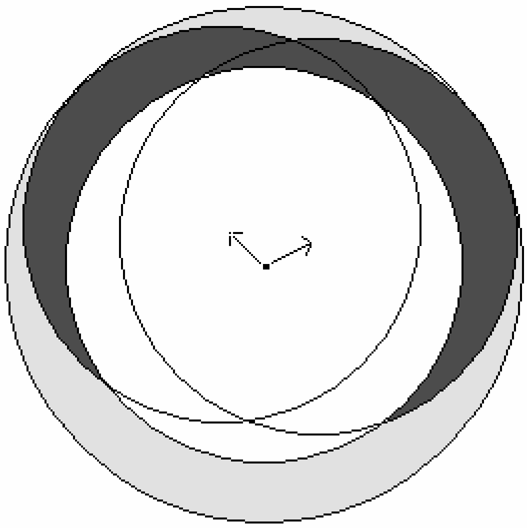

Figure 1 Movement prediction and delay

Given the global Internet as premise the model must take into account delays on a global scale. Such delays are lower bounded by the signal propagation speed, which depends on choice of media. Basically two main choices present themselves [12]: Wired fiber optical giving a lower delay bound of 0.2 second and Low Earth Orbit (LEO) satellite systems with a lower delay bound of 0.13 second. Since LEO systems are bandwidth limited and affected by problems of shared bandwidth their usability for such transmissions is doubtful at best. Even new systems are limited to around 1 Gb/s [13]. To these basic delays must be added delays in active equipment such as switches and routers.

COMPRESSION

There is potentially three methods of compression that may reduce the amount of data in transmission and thus lower the bound; the three methods are non-destructive, destructive and RRI.

Non-destructive Compression

The defining characteristic of non-destructive compression is that the original information content can be restored. This sets a limit to the efficiency of the compression method as a dataset, such as an image, can only be reduced to a shorter representation as long as the shorter representation retains the full information content. The information content of a dataset can be calculated as the entropy of the data [4]. Information theory shows two fundamental approaches to such compression, Huffman compression and Lempel-Ziv compression. For these two the theoretical limit for compression shows that a dataset can be compressed to within one bit of its entropy. For an arbitrary dataset, as is the premise here, the problem of this form of compression is that it can not be assumed that any compression is achievable; if the dataset is already in a representation with no redundancy, no compression is possible. Non-destructive compression can therefore reduce average datarates but not peak.

Destructive Compression

Destructive compression extends beyond this limit by reducing a dataset to a representation with lower entropy. Doing so prevents the full restoration of the dataset as there is no longer a one to one relationship between the new representation and the original dataset. A compression therefore introduces error to the contents. The major advantage of destructive compression is that compression rate is only limited by the usage tolerance for the error introduced. There exists a number of efficient algorithms for destructive compression, where optimizations for different types of data and usages exist. The quality loss varies with the data characteristics, and choice of, for instance, filter in Wavelet transform based compression is significant [7]. When applied to a scenario involving arbitrary content it is necessary to take into account worst case scenarios, where no suitable algorithm is found, or where the signal source is unable to determine in real time either the type of data or the receiver tolerance for the systematic error introduced by compression. This puts into question the applicability of destructive compression for reducing data in a full virtual environment. Nonetheless destructive compression is almost universally used and guidelines achievable compression rates with acceptable quality loss in image transmission are here derived from MPEG2 compression, see Table (5).

Table 5 MPEG2 compression ratios [1]

Profile at Level	Compression ratio
mainP@highL (60Hz)	39.8
highP@highL (60Hz)	31.8
mainP@mainL (30Hz)	19.9
highP@mainL (30Hz)	14.9

Retinally Reconstructed Images

One method employing destructive compression is RRI. It differs from other approaches by employing a control loop to include data on the observer in selectively removing data during compression. By adjusting the detail level in images to the non-uniform resolution capability of the human eye it is possible to compress an image without perceptible loss of quality [6]. Since the human eye only resolves details at highest angular resolution in an area a few degrees across centered on focus, this technique is most effective with images covering a large angular width giving up to two orders of magnitude compression [6]. The technique is dependent on prediction of the position of eye focus on the image; thus, as the control loop latency increases an element of uncertainty is introduced into this prediction and compression must be reduced to avoid potential perceived loss of quality. The loop delay includes transmission delay and processing time for reconstructing the image. Algorithmic complexity for reconstruction is reported as polynomial to the number of image pixels with figures of tens of MFLOPs for 30 frames/s at 256x256 resolution. Given the image size considered above this would place computational requirements at least in the TeraFLOP range. The feasibility of such techniques has been demonstrated for normal eye movement up to a loop delay of 100 ms; beyond this limit and if taking into account saccadic eye movement, where accelerations of 40.000°s^{-2} and peak velocities of 500°s^{-1} are observed, stability of the system has not been demonstrated [8]. This will render the technique ineffective for more than half of all potential connections [12].

THE UPPER BOUND

To calculate the upper bound based on the premises and compression methods it is necessary to consider the least favourable cases; this includes connections spanning the globe, where application of RRI is highly doubtful, and peak datarate in the image transmission, where non-destructive compression has little or no effect. The degrees of freedom in human movement giving rise to uncertainty in position prediction are, as yet, not well defined. Given transmission loop latencies of at least 0.2 second for global connections the moving observer factor is assumed equivalent to a doubling of angular resolution (ar). The image surface (is) is given as either the full sphere or the 160° by 90° section. Colour depth (bpp) is 24 bit and framerate (fr) is given for both 50 and 500 Hz. The figures for destructive compression (dc) are set to a factor 20 based on the MPEG2 values. The bit rate is calculated by Eq. (1)

$$Bitrate = \frac{ar * is * bpp * fr}{dc} \qquad \text{Eq. (1)}$$

Bitrate values for different combinations are given in Table (6).

Table 6 Upper bound, values in bits/second.

Full sphere 160° by 90°		Normal vision	Hyperacuity
With Destructive compression	50 Hz	1.4e+11 5.0e+10	1.4e+13 5.0e+12
	500 Hz	1.4e+12 5.0e+11	1.4e+14 5.0e+13
Without Destructive compression	50 Hz	2.8e+12 1.0e+12	2.8e+14 1.0e+14
	500 Hz	2.8e+13 1.0e+13	2.8e+15 1.0e+15

These values are payload only; for a given combination of transmission technologies, such as IP over Ethernet, transmission overhead must be added. Since overhead is technology specific, it is omitted here.

DISCUSSION

The method presented in this paper defines the necessary components of modelling for an upper bound on content driven growth of network access speed. One element is left out of the model, namely the added demand stemming from purely technical applications such as distributed control and surveillance. In a WAN context these applications are still only emerging, and their long term properties are less well defined. Also, with such applications distributed data processing gives much greater opportunity for reducing the amount of data transmitted.

Applying the figures from Table (6) to the prediction of growth from Nielsens Law a cut off point will be reached no sooner than 45 years from now, and the development could conceivably extend throughout the century.

The figures for network access speed presented here may seem astronomical by today’s standard, but so do today’s specifications for computer equipment by 1970 standards. The figures for datarates represent the logical end point for Internet access speed and a benchmark for evaluating capabilities of network transmission capacity in the long term. If network growth is assumed to halt before this limit is reached, whether hyperacuity is considered or not, it should be supported by positive arguments. The figures presented here present a way marker for resource management in network provisioning warning against dimensioning of capacity based on present day applications and demand.

A major question raised by these figures is whether network technologies will be capable of supporting such datarates. This question not only concerns access speed, but also the transport nets; even if link capacity for the given datarates is feasible this may not hold when traffic is aggregated. Further research into these questions is necessary.

REFERENCES

[1] **ISO/IEC 13818-2 Information technology -- Generic coding of moving pictures and associated audio information: Video** ISO/IEC 1997.

[2] J. Nielsen, "Nielsen's Law of Internet Bandwidth," **Jakob Nielsen's Alertbox for April 5, 1998** http://www.useit.com/alertbox/980405.html

[3] The Swedish ICT Commission, "General guide to a future-proof IT infrastructure" **Report 37/2001, Technical Report,** The Swedish ICT Commission, June 2001 http://www.itkommissionen.se

[4] T.M. Cover, J.A. Thomas, **Elements of Information Theory,** Wiley-Interscience, 1991, ISBN: 0471062596.

[5] J.T. Klosowski et al., "Deep View: High-resolution Reality" **IEEE Computer Graphics and Applications,** vol. 22 (3), 2002, pp. 12-15.

[6] T. Kuyel et al., "Retinally Reconstructed Images: Digital Images having a Resolution Match with the Human Eye" **IEEE Transactions on Systems, Man & Cybernetics, Part A (Systems & Humans)**, Vol. 29 Issue.2 1999

[7] A. Jensen, A. la Cour-Harbo, **Ripples in Mathematics,** Springer Verlag; 2001, ISBN: 3540416625

[8] P. M. Sharkey, D.W. Murray, "Feasibility of Using Eye Tracking to Increase Resolution for Visual Telepresence." **IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation.** Vol. 2. 1997 p. 1078 -1083, ISBN 0780340531

[9] H. Volz, J.M. Zanker, "Hyperacuity for Spatial Localization of Contrast Modulated Patterns" **Vision Research** Vol. 36 Issue 9, 1996.

[10] G. Westheimer, M.W. Pettet, "Contrast and Duration of Exposure Differentially Affect Vernier and Stereoscopic Acuity" **Biological Sciences,** 1990, Vol.241 Issue.1300 pp. 42-46.

[11] J.M. Zanker, J.P. Harris, "On Temporal Hyperacuity in the Human Visual System" **Vision Research** Vol.42, 2002 pp. 2499-2508

[12] T.P. Knudsen, A. Patel, J.M. Pedersen, O.B. Madsen "Signal Propagation Delay as a Barrier to Control Applications in WAN." **The 2004 International Conference on Parallel and Distributed Processing Techniques and Applications Conference Proceedings,** 2004.

[13] J.V. Evans, "The US proposed new multimedia communications satellite systems" **2000 Aerospace Conference Proceedings, IEEE,** 18-25 March 2000.