Network Bandwidth Estimation: A System Dynamics Approach

Anas Tawileh and Steve McIntosh
Cardiff University School of Computer Science
Queen's Buildings, 5 The Parade, Cardiff, CF24 3AA, Wales
+44(0)29 2087 4204/+44(0)29 2087 4598
m.a.tawileh@cs.cf.ac.uk sbm@cs.cf.ac.uk

Abstract

In this paper, we propose a new approach to network bandwidth estimation based on System Dynamics modelling. The paper discusses existing approaches to bandwidth estimation and network capacity planning. Limitations of these approaches are presented and the case for using System Dynamics is made. Applicability of the proposed approach is demonstrated through a real world network planning project for a distributed logistics application. A practical computer simulation model was developed to predict bandwidth requirements for the project's network. This model provides system planners with the ability to test different possible scenarios in order to make informed decisions about the system architecture. We show through practical results of the simulation runs and the insights gained during the process that the System Dynamics approach offers an effective solution to the problem of network bandwidth estimation and system planning. The paper concludes with a review of the results and pointers for further research.

Keywords: Network bandwidth estimation, network planning, System Dynamics.

Introduction

Dependence on computer systems in all aspects of human activity has been growing significantly during the past few years. Computer based applications, interconnected by complex and interdependent networks became the spiral cord of modern commercial, social and cultural communications. Increased demands are placed on Information Technology (IT) infrastructures with the continuous introduction of new technologies and the automation of legacy business processes. In this highly automated environment, network reliability and dependability are essential elements to guarantee the smooth and stable conduct of functional transactions, particularly in mission critical applications. Tolerance for system downtime and excessive delays in response time is shrinking. Building reliable information systems is a primary concern for virtually every organisation.

In order to satisfy these increased reliability and dependability requirements, networks should be planned, designed and implemented with appropriate capacity to accommodate the traffic demands of the applications they will serve. This will require careful analysis of the types of applications using the network, traffic and usage patterns and future expansion possibilities. When properly designed and implemented, computer networks will provide transparent transfer of information between computers and information systems, with minimum delays or disruptions to users' requests. Significant productivity gains could be accrued as a result. In mission critical applications when delays may have disastrous effects, due diligence in designing networks is a prime concern.

Further, in Government and defence areas, the approaches to contracting for logistics and information systems are undergoing change. In the Defence arena, major capital equipments are to be provided by industry, with the Ministry of Defence (MoD) contracting for the provision of specified levels of availability, or with the industrial supplier being contracted to maintain equipment. (Similar approaches to the provision of equipment and services are being increasingly adopted in other areas, such as the provision of railway rolling stock and even school buildings.)

Such initiatives both change the focus of information requirements within the MoD (for example, requiring it to focus on information to manage a service contract, rather than to manage equipment availability) and the flows of information from static and deployed locations, with the industry supplier requiring near real-time serviceability and usage information. Within the UK, the major static and deployed communications infrastructure to support Defence logistics information systems are provided by a major consortium of industrial organisations, and the contractual arrangements are complex. One consequence of this approach to the provision of these vital services is that new IS/IT projects must provide bandwidth estimates prior to the accreditation of the project for hosting on the service provider's infrastructure.

The field of network bandwidth estimation has emerged to provide answers to these increasingly important issues. Its primary aim is to predict the amount of traffic that will cross the network in order to design its capacity accordingly. Several approaches were proposed and many tools developed to aid this process. However, each approach has its limitations. We propose a systematic method to estimate and predict network bandwidth based on System Dynamics.

In this paper, we examine an example of network bandwidth estimation relevant to a real world project. We start by explaining the notion of bandwidth estimation, and describing the project and its requirements. Different approaches to estimate network capacity and bandwidth requirements are then presented and compared. We then propose System Dynamics as an appropriate tool to predict network bandwidth utilisation. The process of developing the System Dynamics model is presented. Lastly, the model is utilised to test several design options and to explore several possible scenarios. The paper concludes with a brief discussion of the implications of this work and the directions for future research.

Background

Network bandwidth is a metric that measures the data rate at which the network path can transfer information (Prasad et al., 2003). This metric has been growing in importance with the rapid technological development and the introduction of new highly demanding applications. Many applications compete for a share of the limited available network bandwidth, and unless planned and designed properly, network users may suffer from slow response times and application timeouts. This will decrease the dependability of the network and may cause significant productivity losses.

Available bandwidth is a critical determining factor in the chosen architectural design of the overall IT system. Bottlenecks (Keshav, 1991) are network connections that have very limited bandwidth available for network traffic. While modern networking technologies have increased the available bandwidth in many local area applications, wide area networks still suffer from extremely limited bandwidth availability. This may not be attributed to inadequate technologies as much as it is a result of the prohibitive costs associated with implementing these technologies on a large scale. Without careful planning to avoid bottlenecks and bandwidth limitations, these low capacity links may impair the operations of the whole system.

In the project that motivated this work, the organisation requested the development of a tool to estimate the network bandwidth requirements of a new application, that was still in the requirements stage, in order to be able to ensure that appropriate capacity was available using the existing industry supplied infrastructure. A contract was placed on a large consulting company who developed a model using MS Excel.

During the presentation of the spreadsheet based model, it became clear that a major failing was the lack of any appreciation of system behaviour over time. The Excel based model was based on a worst-case scenario in which the network operates at full capacity. However, such scenarios are rarely sustained in reality. Moreover, the model also ignored the impact of feedback that will occur as packets drop on the network and re-transmission of information is requested. Accounting for these factors requires consideration of behaviour over time, a feature not readily available in MS Excel. The authors took this as an opportunity to think about the dynamics of a network, and use the problem scenario to investigate both current practice in bandwidth estimation, and the feasibility of using System Dynamics to develop a suitable alternative model. This approach should support the derivation of recommendations that may lead to enhancements in the system's reliability and dependability.

The software application being considered is to be operated in several remote sites, each composed of a local area network and a local server. The local server hosts a subset of the central logistics database in order to enable it to respond to local requests. Each remote site is connected to the main central site through a satellite (VSAT) connection. Due to its high costs, the VSAT connection has only 512 Kbps of network bandwidth. All user requests for information from the application that can not be answered locally should be forwarded to the central site to be processed by the main server.

Depending on the availability of required information, traffic from clients' requests may need to traverse the local network or travel across the satellite link to the main central site. During its journey, traffic is subjected to the availability of networks and servers, and to the limited available bandwidth in each network link. The effects of these factors should be determined in order to adapt the system's design in such a way as to guarantee the required level of service. In the next section we will review current approaches to bandwidth estimation within the context of this project.

Approaches to Bandwidth Estimation

Many tools and techniques have been developed to predict the performance of computer networks. Broadly speaking, all these tools fall under one of two different approaches: discrete event simulation and mathematical modelling (Wombell, 1999). In the first approach, bandwidth utilisation and network performance are estimated by modelling every individual packet traversing the network. On the other hand, the main idea of mathematical modelling lies in building an abstract model of the network consisting of a number of equations founded on a sound theoretical basis. This model is then used to calculate traffic patterns and infer bandwidth utilisation trends and expected network performance (Wombell, 1999).

The NS Network Simulator (McCanne & Floyd) and The Georgia Tech Network Simulator (Riley, 2003) are network simulation tools based on the discrete event approach. Network bandwidth utilisation is estimated using a simulation engine that tracks packet transmissions from origin to destination (Breslau et al., 2000). Some tools incorporate additional features to increase the accuracy of estimations, such as queuing behaviour and parallel processing.

The mathematical modelling approach includes a significant number of tools, such as CACI Network Simulation Suite and NeTraMet (Brownlee, 1997). These tools collect statistical information from the network in order to produce an overall picture of traffic and link utilisation patterns. This can be achieved by either actively probing the network to collect data or by other passive means that rely on different networking protocol characteristics to deduce the required information.

Both approaches have significant limitations that affect their utility in the planning and design phases of computer network projects. For example, almost all the existing bandwidth estimation techniques are based on a single-link model, with constant rates of cross traffic (Jain & Dovrolis, 2004). Moreover, while network traffic and utilisation are inherently dynamic, most of the available tools do not account for dynamic changes in network traffic and consequently the available bandwidth (Prasad et al., 2003) (Jain & Dovrolis, 2004). This is probably the most limiting factor to the usability of these tools in planning and designing networks.

To identify bottlenecks in the network, current bandwidth estimation tools use end-to-end capacity estimation techniques. In this context, a distinction should be made between the narrow link and the tight link. The narrow link is the link with the lowest capacity in the network, while the tight link is the link that has the lowest available bandwidth at a certain moment. Network bottlenecks are determined by the tight link, not the narrow link. While network bandwidth estimation tools can predict the capacity of the narrow link, they are usually incapable of guessing the available bandwidth at the tight link. Furthermore, estimation accuracy of these tools declines significantly when multiple bottlenecks exist in the same network (Jain & Dovrolis, 2004).

Van den Nieuwelaar & Hunt (2004) have also identified user friendliness and lack of consideration of network topology and architecture as major limitations of existing traffic management and network design tools. They conclude that while most of the existing tools can be used for traffic management and measurement, their usefulness in network design is rather limited (Van den Nieuwelaar & Hunt ,2004).

We propose the use of System Dynamics, first introduced by Jay Forrester (1973), to address the limitations we have identified in current network bandwidth estimation approaches and to provide an effective method to aid the capacity planning and design of mission critical networks. The problem of network bandwidth estimation is characterised by complexity, involves many interacting factors and encompasses feedback and delays. Therefore, it can be reasonably argued that bandwidth estimation is a systemic problem, and that approaching the situation with a systems view may yield useful insights. System Dynamics provides a systematic approach to deal with real world systemic problem situations that accounts for the interdependence and interactions between different elements of the system. It also recognises the time delays that usually occur in these systems.

System Dynamics models capture the interrelationships between the system's elements and enable the analysis of causal loops that affect the behaviour of the overall system. Furthermore, by using appropriate computer software, simulation models can be built to experiment with different scenarios and policy options. Such experimentation can provide valuable insights into the system's behaviour and inform the decision making process. In the context of this paper, a simulation model would facilitate the testing of several network design options before the network is physically implemented. It can also weigh different options against each other in order to produce the optimum and most cost effective network utilisation scenario.

User interaction with the simulation model can be facilitated through an elegant graphical user interface that would provide a visual, easy to interpret representation of the traffic patterns within the network. The interface also offers users a flight simulator-like dashboard to change the model's parameters and to immediately observe the impact of these changes. For example, the model user can change the link capacity of a certain network or increase the size of exchanged messages in order to examine the changes in the network traffic behaviour.

Model Development

The aim of this modelling exercise is to produce a simulation model that could assist the organisation in estimating the network bandwidth requirements of its logistics application. This would facilitate the planning, design and implementation of appropriate networks to achieve the required system availability. The logistics application is deployed in a distributed architecture. It consists of a number of remote sites and a central location. Each remote site contains a number of users, a local area network (Remote Site Network) and a local server (Remote Site Server). A VSAT interface unit is attached to the local server to enable it to send traffic across the satellite link.

The logistics application is based on message passing between users and the servers hosting the logistics database. When users need specific information, a request message is sent to the local server in the remote site first. The local server hosts a subset of the main logistics database that contains information that may be needed at this specific site. The server checks the request and searches its local database for any matching entries. If a matching entry is found, the server sends a response message to the requesting user. An update message is also sent to the central server at the main site to add the transaction to the logistics application log. If the request can not be answered locally, the main server forwards the message to the central server through the satellite link. The influence diagram of the interacting factors affecting network bandwidth utilisation is reproduced in Figure 1.

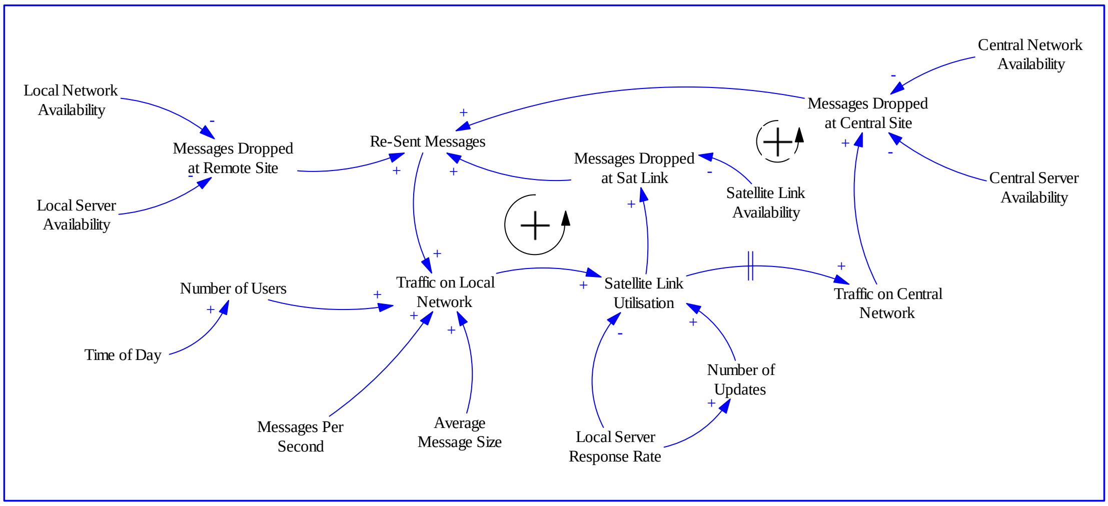

Figure 1: Influence Diagram of Network Bandwidth Utilisation

In order to accommodate the distributed architecture of the system, and to facilitate modifications and flexible changes, the simulation model was developed in a modular structure. This modularity enables the model to adapt to possible architectural changes, and provides decision makers with an easier way to add or move remote sites as required. To achieve this goal, the remote site was modelled as a sub-model in ISEE's iThink software. Figure 2 shows the iThink sub-model structure for the remote site.

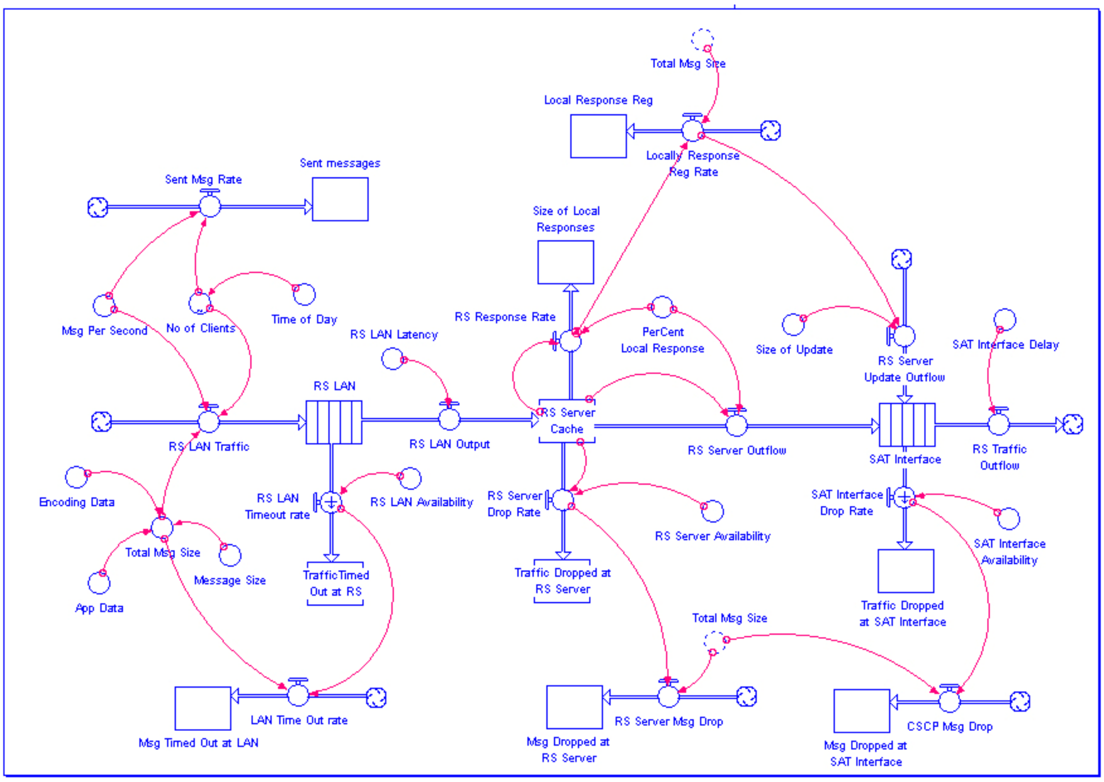

Figure 2: Remote Site Sub-model

Traffic on the local network is determined by the active number of users at any particular point in time (active users are the users using the logistics application). Obviously, the number of active users changes during the day, and consequently affects the network utilisation. The usage pattern at each remote site is defined by a graphical function that changes the number of active users at different times of the day. Network bandwidth utilisation is also affected by the size of each individual message. The message size consists of the data content of the message, the application headers and the encoding data. Each remote site has a Fast Ethernet local area network with 100 Mbps capacity. Traffic traversing the local network incurs losses due to the network availability. The model in Figure 1 keeps track of all traffic and messages dropped at the local network because of availability constrains.

Upon arrival at the local server, request messages are checked against the local database for a matching response. The amount of responses found is determined by the size of the local subset of the logistics database. For each message answered locally, the server sends an update message to the central site. Due to hardware and software limitations, the local server will never be able to process all incoming traffic. Some messages will be dropped at the server as a result. The local server will then send all messages that could not be answered locally to the central site through the SAT interface unit.

While local area networks possess relatively high capacity, the satellite link was a major concern. Satellite links have two serious limitations: firstly, the bandwidth they provide is very low, and secondly, they suffer from significant delays. The satellite link used in this project has only 512 Kbps of capacity. With such low capacity, the link can easily become a bottleneck, affecting the overall reliability and performance of the system. The organisation was concerned that this link may actually prevent the system from achieving the desired reliability requirements, particularly with the increasing number of remote sites connecting back to the central site through the narrow satellite link. In addition, the satellite connection availability will cause some messages to drop while crossing the link.

Messages that traverse the satellite link enter the central site at the SAT Interface unit. Figure 3 shows the structure of the central site sub-model. The central site includes a Fast Ethernet local area network operating at 100 Mbps. Messages cross the central site network to arrive at the central server. This server hosts the central logistics database, which theoretically should contain the information needed by all remote sites. In order to deal with special cases where requests cannot by answered by either the remote server or the central server, the Central Data Base unit was established. This unit basically handles special cases and system errors. The model tracks all dropped messages due to availability of the SAT Interface unit, the central site local area network and the central server.

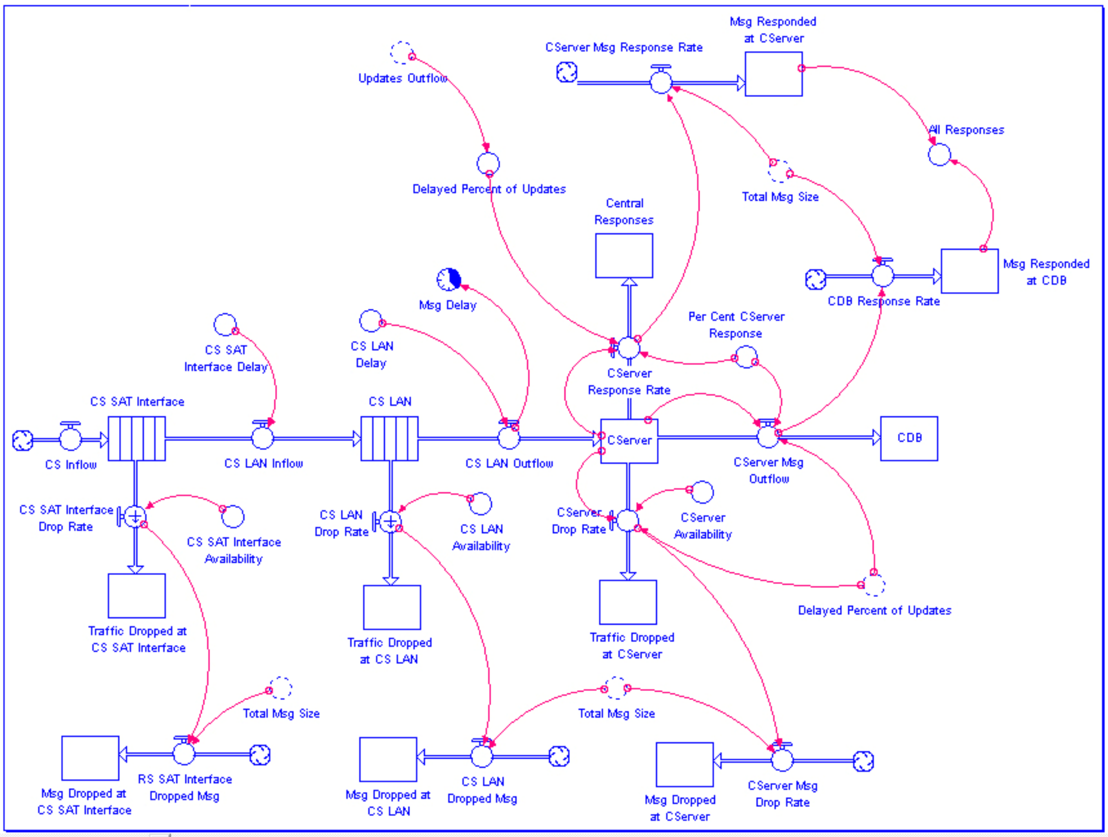

Figure 3: Central Site Sub-model

The complete system model is presented in Figure 4. In order to achieve its intended objectives, the system must provide reliable support for the logistics function in the organisation. System users should be able to get the information they need in a timely manner. It should be possible also for managers to establish a certain level of confidence in the system and its availability and response time. This requirement translates into minimising delays in response time and avoiding bottlenecks and traffic congestion at critical links. Two major indicators provide insight into these aspects of system behaviour: the total delay (the time it takes each message to arrive at the central server) and the utilisation of the satellite link. By observing these two factors in different usage scenarios, managers can get a clearer idea about the behaviour of the system and the factors that have the greatest leverage in changing this behaviour.

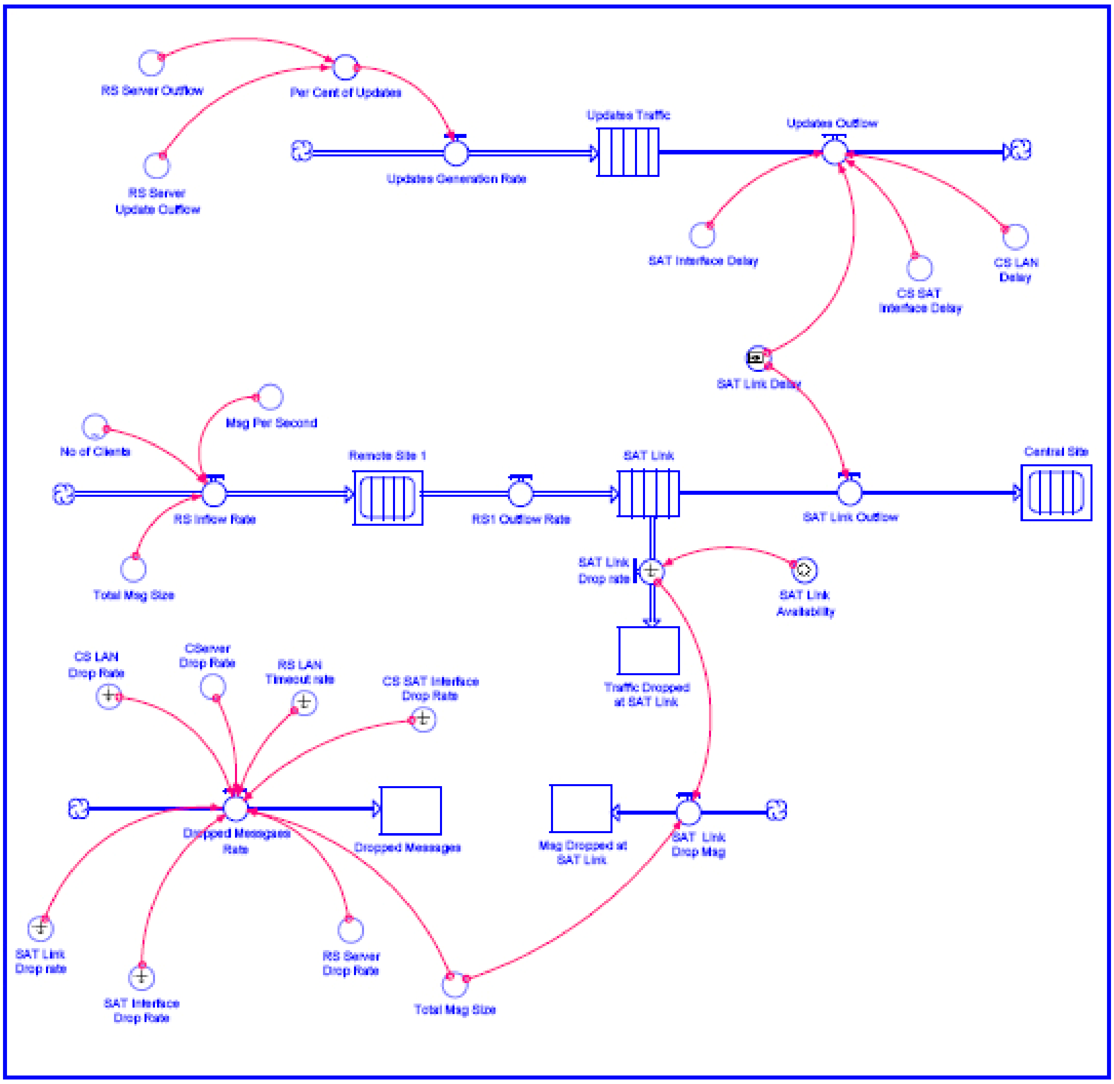

Figure 4: Complete Bandwidth Estimation Model

Ease of use has been reported as a serious limitation of previous network bandwidth estimation tools (Van den Nieuwelaar & Hunt, 2004). Because of the large number of interrelated factors involved, interacting with the model at this level may become an arduous task, especially for non-technical audiences. A high level user interface was developed to facilitate interaction with the model and mask the complexity of its structure. The user interface is reproduced in Figure 5.

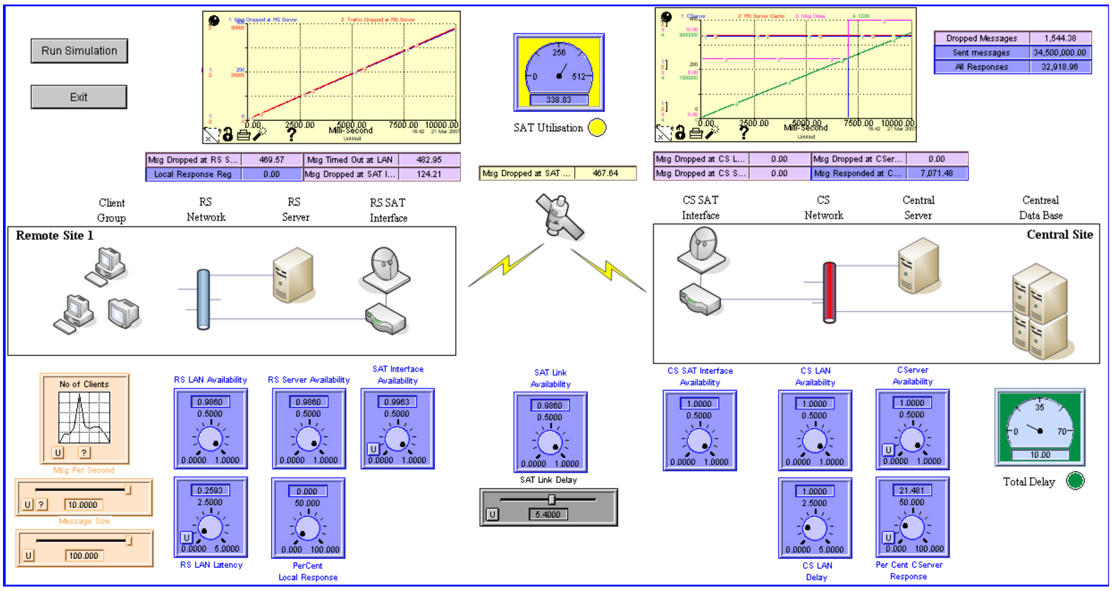

Figure 5: User Interface for the Bandwidth Model

Model Validation

Before the developed model can be utilised to experiment with different possible scenarios, an acceptable level of confidence should be established in its behaviour. The validation process was conducted in consultation with subject matter experts with significant knowledge about network design and bandwidth estimation. Several simulations were run with different sets of input parameters where the system behaviour could be predicted. In addition, a simulation run was conducted with input parameters that could be reproduced in a lab environment with a single remote site connecting to the central site (Table 1 includes the values used in this simulation). The results produced by these simulations were highly comparable to expected outputs, which confirm the validity of the model and its applicability to the project under study.

Input Parameters			
Messages per Second	1	SAT Link Availability	98.20 %
RS LAN Availability	99.00 %	CS LAN Availability	99.90 %
RS Server Availability	98.00 %	CServer Availability	99.98 %
Local Response	80 %	Message Size	10 KB
Simulation Period	1 Hour	Application Data	1 KB
No. of Users (Varies with time)	20 max	Encoding Data	1 KB
Results			
Max. Bandwidth Utilisation of SAT Link	373 Kbps	Message Delay	11 ms

Table 1: Base Case Simulation Inputs and Results

The validation simulation results confirm the organisation's concerns of the limited bandwidth of the satellite link. Despite the fact that only one remote site is connected in this scenario, and the relatively low number of users interacting with the system, the traffic generated by this single site consumes more than 70% of the link's available bandwidth.

After the model's behaviour has been validated, it can be utilised to test several possible design and architecture options. The following section discusses the exploration of different alternatives to optimise the usage of available bandwidth in order to achieve the required level of reliability.

Model Exploitation

In order to alleviate the organisation's concerns about the limited capacity of the satellite link, two options can be considered. Firstly, the satellite link may be upgraded to a higher capacity. However, due to the market dynamics of satellite connectivity, this option would incur significant ongoing costs. Secondly, the architecture of the system could be adapted to optimise the utilisation of available bandwidth. This option will also require considerable investments. Therefore, the two options should be carefully weighed in order to make the most appropriate decision.

Several scenarios were developed to experiment with different possible options. In the first scenario, the impact of update messages was evaluated. Two simulations were run with the update message enabled in the first and disabled in the second. Figure 6 illustrates the satellite link utilisation in each case. The graph clearly shows the significant bandwidth consumption caused by update messages. Update messages constitute almost 25 % of the traffic traversing the satellite link. This situation can be improved by designing the logistics system in a way that minimises update messages, such as sending the updates periodically in a batch file. Another solution would be decreasing the size of individual update messages.

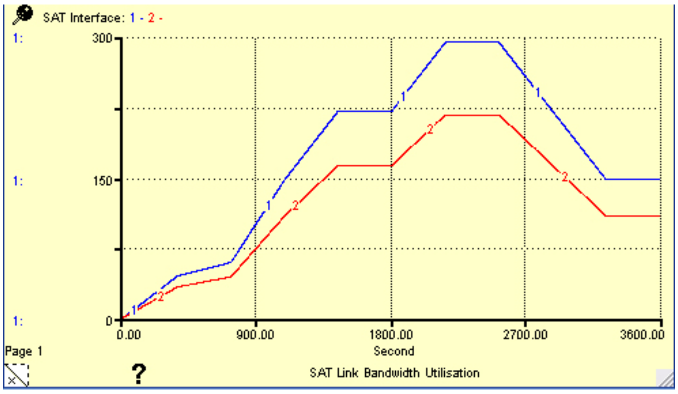

Figure 6: Satellite Link Utilisation (1) with and (2) without update messages.

Bandwidth utilisation in the system is also affected by the number of dropped and timed out messages along the network path. When a request message is dropped, the application's user interface generates an error message, most likely causing the user to resubmit the request again. This would wastefully increase the traffic on the network links. Reducing the amount of dropped messages requires significant investments in hardware and software to increase the availability of network links and servers. Another option would be increasing the capacity of the tight links in order to accommodate wasted traffic. Before committing any investments in improving the system's performance, the organisation can weigh the return on investment of each option and decide accordingly. The graphs in Figure 7 represent the satellite link utilisation and the total number of dropped messages for three scenarios. The input values for each scenario are presented in Table 2. As Figure 6 shows, while investments in availability have substantial impact on the number of dropped messages, its influence on the bottleneck bandwidth utilisation is marginal.

Scenario	*RS LAN Availability*	*RS Server Availability*
1	99.80%	98.00%
2	98.00%	99.80%
3	99.80%	99.80%

Table 2: Input Values for Availability Scenarios

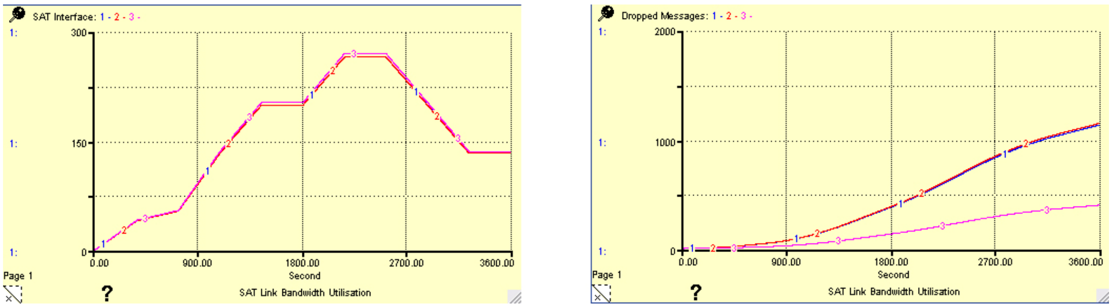

Figure 7: Satellite Link Utilisation and Dropped Messages for Different RS Availability Values

The results of all simulation scenarios clearly indicate the correlation between the number of active users and bandwidth utilisation of the satellite link. The number of users changes with time, and this change is reflected in the bandwidth consumption in the satellite link. Therefore, an option that may be investigated is to purchase satellite link bandwidth on an on-demand basis. In this way, the organisation does not have to expend steep periodic subscription payments to acquire high bandwidth that would be partially utilised most of the time. Before making any bandwidth upgrade decisions, a careful analysis should be carried out to decide on the most realistic active user distribution during the day. This would lead to a much more accurate estimation of the dynamic bandwidth requirements for the satellite link. Consequently, investment decisions can be made in such a way as to maximise the gain in system performance. These conclusions are very difficult to observe in the previously proposed conventional approaches as they emerge as a direct consequence of the complex interactions among contributing factors and the system's behaviour over time.

The scenarios presented in this paper are only indicative of the types of planning experimentation and policy testing that could be performed with the aid of the simulation model. The model can provide an invaluable tool to evaluate the effects of adding additional remote sites to the satellite link, creating a regional site that acts as a hub for several remote sites and to assess the viability of deploying additional applications other than the logistics application on the currently existing infrastructure. These changes may alter the system's behaviour by affecting the traffic patterns. Creating a regional site for example may reduce traffic load on the satellite link and enable connectivity to remote sites through higher bandwidth and lower costs networking technologies. The ease of use of the user interface, coupled with the modular structure of the model itself, facilitates user interaction with the model and reduces the learning curve.

Conclusions

With the growing reliance on interconnected computer systems in everyday life, planning and building reliable networks that can accommodate the increased bandwidth requirements of new applications became critical ingredients for success. Several approaches were proposed and many tools developed to tackle the issue of network bandwidth estimation. However, the existing tools and techniques have some limitations that seriously impede their usefulness in planning and designing new networking projects.

In this paper, we acknowledge the systemic nature of the network bandwidth estimation problem and propose the use of System Dynamics to address network planning issues. A real world bandwidth estimation project was used to provide a scenario to demonstrate the applicability of System Dynamics and aid the modelling process. The project provided invaluable input into the development of a computer simulation model that was utilised to conduct several design options and scenarios. The experimentation process and the resulting insights proved the usefulness of System Dynamics in tackling the problem of network bandwidth estimation.

Further work should be done to integrate the simulation model with network probes that collect real time performance and bandwidth utilisation information from the live network. By analysing collected information and comparing it with the model's outputs, better calibration of the model's variables can be achieved. In addition, unexpected behaviour can be detected and investigated in order to identify its causes and explore possible solutions to rectify them. The model can also be elaborated to include Quality of Service (QoS) provisions for different types of traffic and service levels. By implementing QoS, certain message types or certain users' traffic can be granted higher priority while crossing the network. This would have significant consequences on the overall system performance, and extending the model to account for these changes will enhance its usefulness.

References

Breslau, L., Estrin, D., Fall, K., Floyd, S., Heidemann, J., Helmy, A., Huang, P., McCanne, S., Varadhan, K., Xu, Y. & Yu, H. 2000. Advances in Network Simulation. *IEEEComputer, 33(5).*

Brownlee, N. 1997. Traffic Flow Measurement: Experiences with NeTraMet, RFC2123, http://rfc.net/rfc2123.html, accessed 20 March 2007.

CACI Network Simulation Suite. http://www.caciasl.com, accessed 20 March 2007.

Forrester, J. W. 1973. *Industrial Dynamics*. MIT Press: Cambridge, MA.

Jain, M. & Dovrolis, C. 2004. Ten Fallacies and Pitfalls on End-to-End Available

Bandwidth Estimation. *ACM IMC, October 2004*.

Keshav, S. 1991. A Control-Theoretic Approach to Flow Control. In Proceedings of ACM SIGCOMM 1991.

McCanne, S. & Floyd, S. UCB/LBNL/VINT Network Simulator – NS (version 2), http://www.isi.edu/nsnam/ns/, accessed 20 March 2007.

Prasad, R., Murray, M., Dovrolis, C. and Claffy, K. 2003. Bandwidth Estimation: Metrics, Measurement Techniques, and Tools. *IEEE Network, November/December 2003.*

Riley, G. F. 2003. The Georgia Tech Network Simulator. In Proceedings of the ACM SIGCOMM Workshop on Models, Methods and Tools for Reproducible Network Research (Karlsruhe, Germany, August 25 - 27, 2003).

Van den Nieuwelaar, M. & Hunt, R. 2004. Real-time Carrier Network Traffic Measurement, Visualisation and Topology Modelling. *Computer Communications Volume 27, Issue 1*.

Wembell, S. 1999. Network Capacity Planning. *IEE Review, March 1999*.