Table of Contents

Go Back

Modelling Capacity Requirements for BBC IT Storage Area Network: Experience and Research

Workneh Hailegiorgis[1] and Ddembe Williams[2]
Faculty of Business, Computing and Information Management
London South Bank University
103 Borough Road
London SE1 OAA, UK

[1] e-mail: workneh.hailegiorgis@bbc.co.uk

[2] Tel: +44 (020) 7815 7460 Fax: Tel: +44 (020) 7815 7499
e-mail: d.williams@sbu.ac.uk

Abstract

The research carried out in this paper provides BBC Technology Ltd, which is the current IT services provider to the BBC with a system dynamics model as a decision support tool in storage services management. BBC Technology Ltd recently installed a SAN (Storage Area Network) which is a highly scalable data storage infrastructure to satisfy an increasing level of demand by its IT user base. The System Dynamics model developed in this study is used to facilitate the understanding of SAN capacity utilisation trends and strategic acquisition planning decisions. The model creates learning environment that enables efficient and effective management of data storage capacity requirements in different planning time frames. The results have significant implications for long-term capacity investment decisions for IT Service managers and capacity planning managers.

Key words: Storage Area Network (SAN), System Dynamics, Data Storage Requirements, Modelling, IT Service Management, Decision Support Tool

1. Introduction

The BBC has over 34,000 IT Application users worldwide. It has one of the largest IT infrastructure user base in Europe. This paper is based on the recent acquisition by BBC Technology Ltd of the Storage Area Network (SAN) infrastructure as a step change in storage strategy after experiencing year on year exponential increase of storage utilization. As a media broadcasting organization the demand for electronic data storage has been highly evident as the corporation evolved to integrate the emerging digital broadcast technologies which resulted in the introduction of media rich applications generating large volumes of multi media data across the corporation.

Advances in computing technology have seen the corporate demand for storage capacity requirements grow exponentially over the past decade and this is expected to rise at an even steeper trajectory in the next few years (Dataquest 2000). Disk storage requirement have been rising dramatically enabled by the increasing speed of processing power, and the decreasing cost of computer hardware and support. Business confidence in the electronic data storage environment has increased as corporate data backup and retrieval solutions have matured over this period, providing a high availability platform for data sharing in rich media applications and e-commerce. Over this period also, successive IT departments that provided the service and support for the organisation's information systems infrastructure have not been able to address the problem fully or to provide a strategic solution to the escalation of storage demand. In an attempt to contend with the problem in the short term, different policies and practices have been implemented to fulfil operational service requirements but none have attempted to fully understand how data is generated, used and stored in the organisation, and as a result, have not been able to provide a sustainable solution to deal with the fast growing demand for storage.

After a number of years of escalating storage demand the management of BBC Technology Ltd. recognised that the continuing growth of users' data required a step change in storage capacity planning and have embarked on providing an advanced 'state of the art' solution to the capacity requirement problem using Storage Area Networks (SAN) technology. As part of this process, management have been forced to address the issues of storage scalability, consolidation, centralisation, the administration cost and availability of highly valuable data assets (Leman Brothers, 2001).It is therefore important to fully analyse the strategic capacity requirements of this new platform from both business and technical perspectives so as to enable managers to accurately project and plan for future capacity expansions resulting from data growth and associated demand for storage.

2. Information Technology Capacity Management Environment

2.1 Capacity management process overview

Storage Capacity Management is an unstructured activity. Kleijnen (1980) points out that although the technical performance and storage utilisation measurements using formula timing, simulation or other quantitative techniques are highly structured, the economic evaluation of storage capacity is an unstructured activity. In today's environment predicting storage utilisation and technological development over the next five years as the system dynamics models attempts to do is a complex problem.

The IT Service Management forum using the ITIL (IT Infrastructure Library) framework defines Capacity Management essentially a balancing act that is

(Stewart, 1990) *Cost against Capacity* , the need to ensure that processing Capacity that is purchased is not only cost justifiable in terms of business needs, but also the need to make the most efficient use of those resources. And, *Supply against Demand*, making sure that available supply of IT infrastructure resources matches the demand made on it by the business, both now and in the future, it may also be necessary to manage or influence the demand for a particular resource.

(Stevens, 1980) indicates that the service objectives of timeliness, accuracy, cost and reliability are important in the evaluation of IT storage capacity management. He points out that the goal is not to achieve 100% utilisation, but rather an acceptable level of service such that the IT infrastructure is meeting the need of the users.

2.2 Current practices in capacity management

Capacity is usually one of the key considerations in defining strategic objectives in Information Systems/Information Technology (IS/IT) management. There are a number of mathematical models that are used to define and resolve capacity planning problems. This paper examines models used in addressing data storage capacity planning and forecasting issues in the management decision-making process.

The BBC Storage Area Network infrastructure is managed by BBC Technology Ltd, currently most of the capacity planning and management functions are carried out through the manipulation of Microsoft Excel Spreadsheet. Utilisation data from servers which are attached to the SAN are collected on a fortnightly basis and a full summary is complied every month. The collected data is used for primarily two reasons. The first reason is to monitor and control storage utilisation and availability trends on the SAN infrastructure and secondly to formulate capacity requirements plans based on the trends observed in the operational environment. For the planning process mangers apply formal estimation tools available in Microsoft Excel such regression and time series analysis.

The in-built functions within Microsoft Excel are the primary tools currently used in capacity planning process of BBC Technology Ltd. The application is used to project values and create trend lines based on current storage utilisation data recorded from the SAN file servers. To extend complex and nonlinear utilisation data sets Microsoft Excel uses regression analysis, which is a form of statistical analysis used for forecasting. (Microsoft Excel 2002) Regression analysis estimates the relationship between variables so that a given variable can be predicted from on or more other variables. The worksheet functions used in capacity requirement calculations are *forecast, trend,* and *series* functions.

The *forecast* function calculates, or predicts a future value by using existing values. The predicted value is a *y-value* for a given *x-value*. The know values are existing *x-value* and *y-value*, and the new value is predicted by using linear regression. The known *x-value* is actual capacity utilisation and known *y-value* is time interval. This function is used to predict future capacity requirements based on current levels of utilisation.

The *trend* function returns values along a linear trend. The function fits a straight line using the method of least squares to the arrays of *know_y's* and *know_x's*. Returns the *y-values* along that line for the array of *new_x's* that is specified.

know_y's : is the set of y-values we already know in the relationship y= mx +b
know_x's : is the optional set of x-values that we may already know in the relationship y=mx+b
new_x's: are new x-values for which we want *trend* to return corresponding *y-*values

The *series* function is used by the capacity planner when manual control is required to generate a linear or growth trend using the computer to fill the projection values. In a linear series, the starting values are applied to the least-squares algorithm (y=mx+b) to generate the series. In growth series, starting values are applied to the exponential curve algorithm (y = b * m ^x) to generate the series. In either case the step values or the difference between the first and next value in the series are ignored.

The storage management team in BBC Technology Ltd understand that the spreadsheet functions applied on capacity utilisation data offers limited modelling capabilities and require a lot of time to generate new scenarios and therefore, decisions. They are looking for methods that generate optimal decisions directly without the need for mangers to guess at optimal solutions and analyse those decisions via spreadsheet manipulations. They also require a system to respond quickly to changes and utilise less time for analysis.

3. Capacity Management Modelling with System Dynamics

The linear functions discussed using spreadsheet manipulations are part of mathematical planning models that uses optimisation techniques to formulate decisions in capacity planning. This type of decision-making is also referred to as deterministic as it provides specific values given a set of conditions with a set of objectives. It will input a set of parameters, which are pre-determined over a planning horizon to output an optimised solution to the decision maker. The assumption of "linearity" means that each causal factor impacts the "effect" by a fixed, proportional magnitude. (Richmond 2001) *Linear functions could therefore be defined as a one way view of the problem domain"*. The analysis used in optimisation methods is static in nature with no feedback factors that are causing changes into the system as long as the factors are satisfying the set of objectives.

It is commonly recognised that the power of linear models is limited to explaining past behaviour, or to predict future trends given that, there will be no significant change in the pattern of behaviour that was observed. (Daellenbach 1999) These models don't fully understand outcomes which are unintended, unpredicted and may partially or wholly negate the sought after benefits of a particular decision.

Linear functions in spreadsheet manipulation have advanced with the increase in computational power built on mathematical foundations and it is used extensively in cost and price analysis in capacity planning and management. However it has major short falls, in that it is not able to model emergent behaviours, uncertainty and feedbacks without the need of numerous assumptions, approximations and post model analysis. The linearity conditions imply that non-linear effects such as economies of scale and reliability issues could not be modelled accurately. Linear functions use considerable computer resources to satisfy the large number of constraints and variables in capacity planning. The complexity and size limits the efficiency in providing an optimal solution. .

This paper purposes to apply System Dynamics modelling approach to solve the capacity requirements planning problems identified in BBC storage service management. In particular it will apply the Dynamic Synthesis Methodology as the framework for analysis, model building and simulation that will be discussed in this paper. (Williams 2002) Dynamic Synthesis Methodology (DSM) refers to the integration of theoretical concepts and structuring of parts and elements of a process over time in such a manner to form a formal functional entity by synthesis as philosophy of science. Synthesis is an attempt to fuse the findings of various branches of science into coherent view, in order to explain why things operate the way they do.

Using DSM case study research method is combined with system dynamics modelling to form a framework for aiding quantitative explanation and prediction of the behaviour of complex systems under investigation. "*The advantage of modelling in system dynamics pseudo code is its iterative nature that yields understanding which forms the basis for further analysis, theory testing and extension*" (Williams 2002). (Meadows 1982) The system dynamic modelling approach enhances theory development using case study research method by combining systems paradigms to natural science paradigms.

(Williams 2002) Case study research method is an empirical investigation that probes and examines responses of convenient influences within the operational environment of the task, users and system. System Dynamic modelling approach in conjunction with case study research method will provide the qualitative information that will be used to understand the problem domain in more detail. As the DSM combines the case study and simulation research methods it becomes a powerful tool for problem solving and analysis. Case studies on their own are used to capture a description of real situation while simulation experiments are used to build abstraction of the real world and test the abstraction with formal data analysis. These methods are complimentary which makes DSM a powerful research tool in Operation Research and Management Science environment.

The DSM includes six iterative research process phases, namely: Problem Statement, Field Studies, SD Model Building, Case Studies, Simulation Experiments and Model Use and Theory Extension. The process diagram that outlines this iterative research process, which is used to conduct the research study, is illustrated on Figure 1 overleaf.

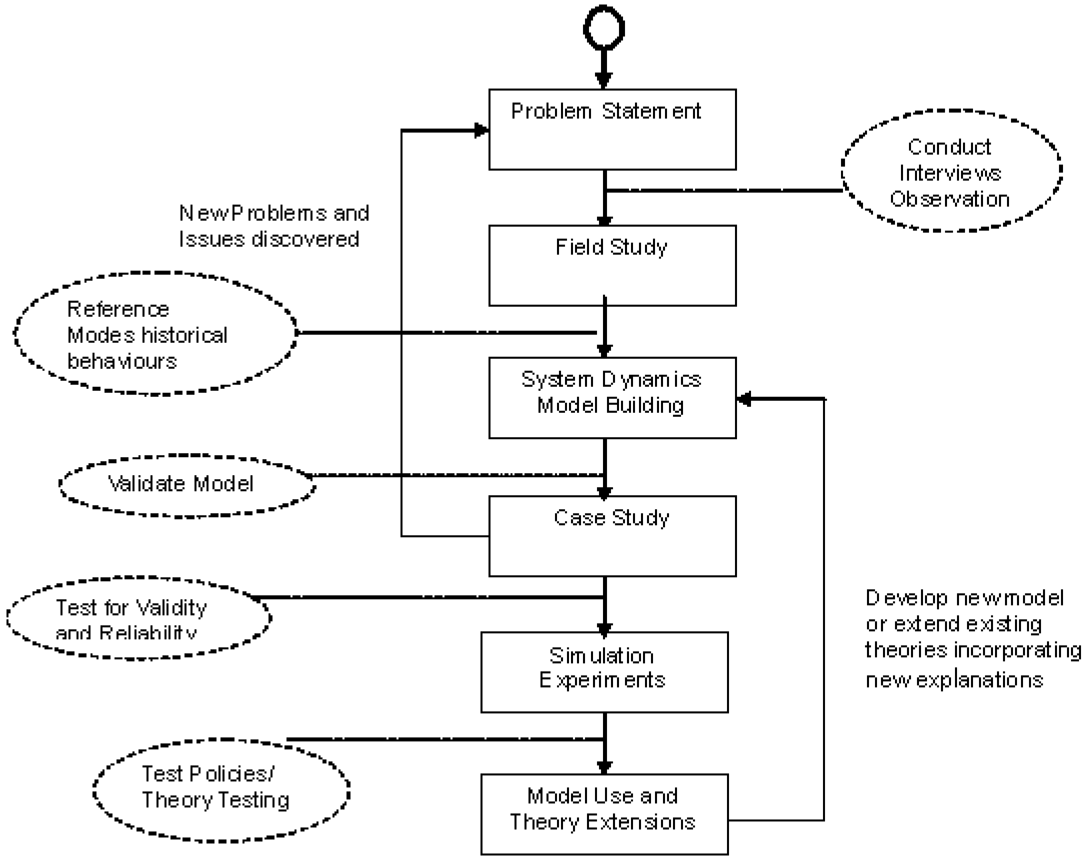


Figure 1 The Dynamic Synthesis Methodology (Williams 2002)

The usefulness of SD for developing and maintaining capacity requirements process modelling and analysis lies in the following (William, Hall and Kennedy, 2000; Williams, 2001)

- System Dynamics SAN Capacity Requirements decision making process model permit the modelling of complex processes with various levels of granularity; that enables the concentration of required level of details without being overwhelmed by lower level of details.

- The SAN Capacity Requirements decision-making process model developed can be used by stakeholders in training situation as one can simulate the process to gain better understanding.

- The use of system thinking and application of SD modelling notation gives the operational managers with a balanced perspective between "hard" (Davis and Vick 1977, Boehm 1981) and "soft" (Checkland and Scholes 1990) systems problem solving paradigms.

IS/IT operational management are now more highly socio-technical processes which need approaches or problem-solving paradigms that can captured both the qualitative and quantitative issues commonly found in complex systems. (William and Kennedy 1997) Paradigms that support both hard and soft issues should be welcomed in operations management.

4.0 Dynamic Hypothesis of the Capacity Management Process

The BBC's storage requirements have been growing at an exponential rate over the past decade. As business requirements for desktop applications started to increase, the data growth rate has been increasing at over 50% -70 % on a year on year basis (EMC Final ITT Commercial Response, 2001). The increase of applications used on the desktop has further intensified as the BBC delivers its digital broadcasting output through the integration of PC based machines running media rich applications for internet and broadcast services. As a result of the ever-increasing demand by the business, BBCT had to make a step change in introducing Storage Area Networks to satisfy the data growth rate by increasing operating capacity using scalable and centralised storage infrastructure.

As more data is generated and a proportion of total storage capacity becomes operational, the available capacity for data storage goes down. As the rate of growth in data stored on the BBCT storage infrastructure increases, the available capacity provided is utilised at a faster rate and becomes operating capacity. As available capacity decreases the operating capacity increases, the live data storage environment BBCT needs to support and maintain. The reduction of available capacity for storage means the reduction in data growth rate and increase in service level violations as utilised operating capacity exceeds the agreed service levels of availability. The increase in service level violations is an indictor of the interruption of service continuity and the reduction in data production that is vital to the organisation's business operations. It is the reduction in data production that has a slowing effect on data growth rate until available capacity has been increased to a level where the production of data can resume at the rate required by the business. The relationship between data growth rate, available capacity, operating capacity, service level violations and data production is a relationship that has been driving the level of total storage capacity at an exponential growth rate and this is the central problem that is under investigation in this dynamic hypothesis.

4.1 Problem Statement

In the drive to meet business requirements and provide the agreed level of service quality, BBCT is under constant pressure to balance the delivery of cost effective IT solutions while maintaining an expected level of service quality to its customer base. However, with a scaleable capacity expansion technology now in place using EMC's Storage Area Networks the cost of capacity acquisition will also escalate dramatically if capacity requirements are not planned effectively and demand for storage with resulting data growth rate is not fully monitored and controlled. The characteristic of the problem indicates a vicious circle of capacity requirements with increasing levels of data storage and acquisition cost at every round of the planning process. This balances the demand from the users to produce and store exponentially increasing data assets. When this is viewed against BBCTs' obligation underpinning its relationship with the BBC as its main customer base, reducing IT Service costs looks difficult to achieve with escalating storage capacity cost driven by increasing data growth rate and growth rate projections. The challenge for BBCT is to provide quality storage service to satisfy users' storage capacity requirements and at the same time enable the optimum utilisation of acquired operating capacity cost effectively. The dynamics described in this hypothesis are used to define the relevant system variables in the capacity requirements planning model.

4.2 Defining Key Variables

The problem statement defined above suggests that the system dynamics Capacity Requirement Planning model could be characterised by the fluctuations in users' desktop application data production, available storage capacity and service level violations resulting from the operating capacity provided by BBCT to the customer base and the rate at which this storage environment is utilised for data production. These important variables create the framework for the quantitative planning process in relation to capacity requirements. With the aid of simulation techniques it will be possible to understand the effect of the changes in the variables over time and to make the optimal decision that is able to balance cost against capacity and supply against anticipated demand. The following key variables identified in the problem statement and confirmed by BBC field observations in IT service operations are important to explain and describe the changes in the capacity requirements over different time frames. These key variables are as follows {units of measure or dimensions are given in curly brackets};

BBCT Customers: The number of customers signed under the BBCT service level agreement that have network storage requirements and are able to generate and store data on BBCT SAN infrastructures {Users}

Desktop Application Data Production: Relates to the total amount of data generated by BBCT customers connected to the server types hosted on the SAN. These server types are currently Personal Data Servers (FS), Shared Data Servers (RD) , E-mail/Public Folder Servers (XU). {Megabytes}

Data Growth Rate: Is the rate at which the data type stored on the SAN increases over a given period of time. {Megabyte/year}

Operating SAN Capacity: This is the amount of storage space that BBCT customers are utilising, and have already populated with data types. This includes capacity allocation through storage mirrors and resilience building storage configurations. {Megabyte}

Available SAN Capacity: The amount of storage space that is provided under the service level agreement for the utilisation of BBCT desktop system users, also known as headroom capacity {Megabyte}

Total SAN Capacity: This relates to the total amount of SAN capacity that is fully owned and managed by BBCT which is the total sum of Operating SAN Capacity and Available SAN Capacity. {Megabyte}

Service Level Violations: is the total amount of storage capacity that is utilised over and above the threshold of the agreed service level of storage availability for the data types under investigation. Or it could represent the quality of service provided by BBCT under capacity levels. {Megabytes}

Additional Capacity Acquired: relates to the total amount of storage capacity acquisition derived from requirements of new businesses and expansion of the operational environment in maintaining the agreed level of available capacity. {Megabytes}

Capacity Required for New Business: The capacity that will be allocated to new services that will be hosted on the SAN in consideration of available capacity and data growth rate of the service. {Megabytes}

Operational Capacity Expansion Required: The capacity requirements derived from operational data storage scarcity as indicated by service level violations and deficiencies in quality of services. {Megabytes}

4.3 Reference Mode Behaviours

Reference modes can provide an insight into the dynamics that is present in a system under research. The inherent problem facing SAN capacity requirement planning is the exponential data growth rate with high trajectory, the absence of policy to optimise the operating SAN capacity, inability of the service provider to influence the demand for storage under service level agreements and the escalating total cost of storage ownership. The reference modes define the behaviour for the key system variables under review and give an insight into their attributes, relationships and interaction in an attempt to fully explore the problem area. The understanding gained can be used to hypothesize possible effects of leverage applied on key variables such as operating SAN capacity , data growth rate , desktop application data production and additional capacity acquisitions. Figure 2 indicate the effect of data growth rate on available SAN capacity , applications data production and service level violations as the number of BBC Technology Ltd (BBCT) customer base expands over time.

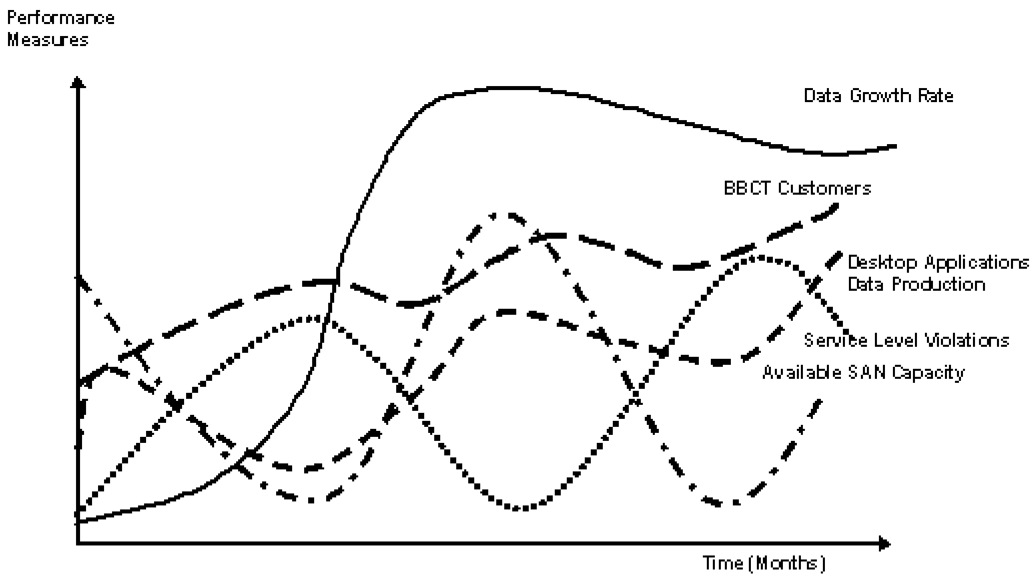


Figure 2 Reference modes of Key Variable

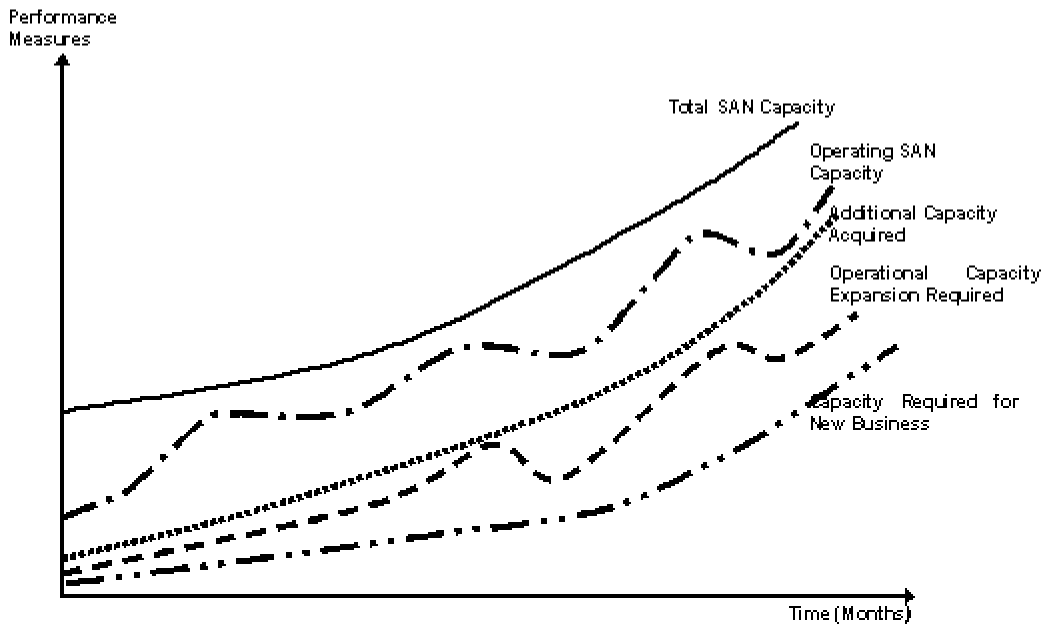


Figure 3 Reference modes of Key Variable

Figure 3 above indicates the effect of operating SAN capacity as BBCT management continuously acquire more storage capacity to fix the availability problems and attempt short term optimisation exercise on current operating capacity. The net effect is an increasing total SAN capacity, operational and new business capacity requirements and capacity acquisitions. (Manni & Cavana 2000) The reference mode indicated by the operating SAN capacity and service level violations can fit into three identifiable system dynamics archetypes 'fixes that fail', 'unintended consequences of expediting' or 'shifting the burden' as a result of a continuous application of short term fixes only for the problem to reappear in greater intensity which is characterised by the exponential increase of additional capacity acquisitions and total capacity over time. The reference modes indicated above provide a vital understanding into the underlying dynamics present in the system.

The relationship between the key system variables in the SAN capacity requirement model could be demonstrated using the causal loop diagram on Figure 1.2 as a basis for constructing a dynamic decision support system. The diagram provides a platform for research into capacity requirement planning success and provides the key propositions that can be tested using data from field studies and validated by stakeholders in SAN capacity. Figure 1.2 provides the theoretical framework of the underlying hypothesis proposed in this paper. It is the relationship and interaction of variables as illustrated in the diagram using cause and effect analysis that determine the behaviour patterns over time as a result of effects from improvement in planning capacity requirements, optimised capacity utilisation and cost effective acquisitions. The initial capacity requirements planning feedback structure contain eight dominant feedback loops of which two are Reinforcing loops (R) and the other six are Balancing loops (B).

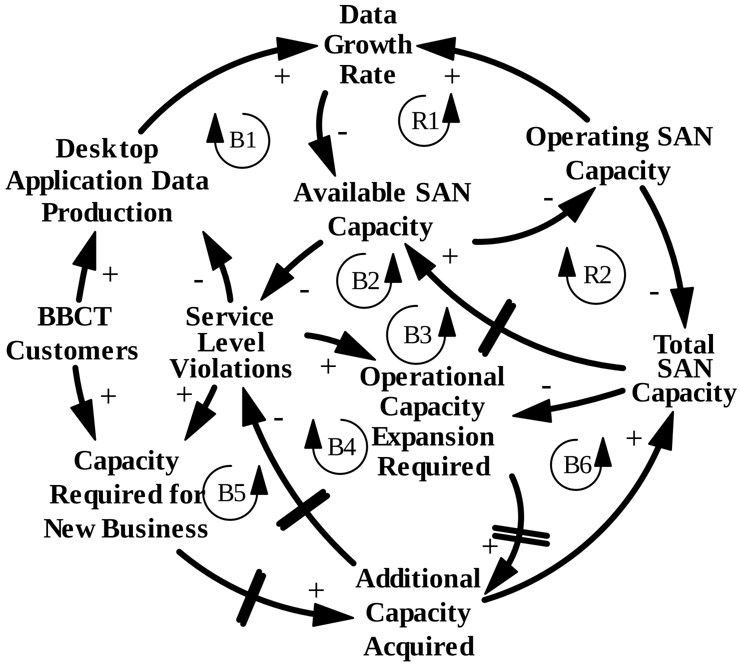


Figure 4 Dynamic Hypothesis of the SAN Capacity Requirement

The interactions amongst the variables identified in the diagram determine the capacity requirements levels on SAN. As users' desktop data production increases, the data growth rate levels accelerate, which reduces the available SAN capacity at a much faster rate than anticipated, affecting service levels and quality which will result in further capacity acquisitions. These increase total SAN capacity, hence headroom available capacity and the cost to BBCT. As the available capacity gets utilised the operating capacity also increases which requires support and maintenance, increasing the total cost of ownership to BBCT. As operating SAN capacity increase data growth also increases, requiring further increase in availability. These continuous dynamics for capacity requirements are explicitly illustrated in Figure 5.3 as data growth rate –capacity availability loops (R1, B2 and B3): customer storage demand – capacity supply loops (R2, B1, and B6): capacity acquisition – service quality (B4, B5).

5.0 The Conceptual Framework for model design

The conceptual framework design covers the main areas of the storage capacity requirement process that has been observed in the field study. This will include variables, values, factors in the relationship between customer and service provider that will be vital to include in the model development.

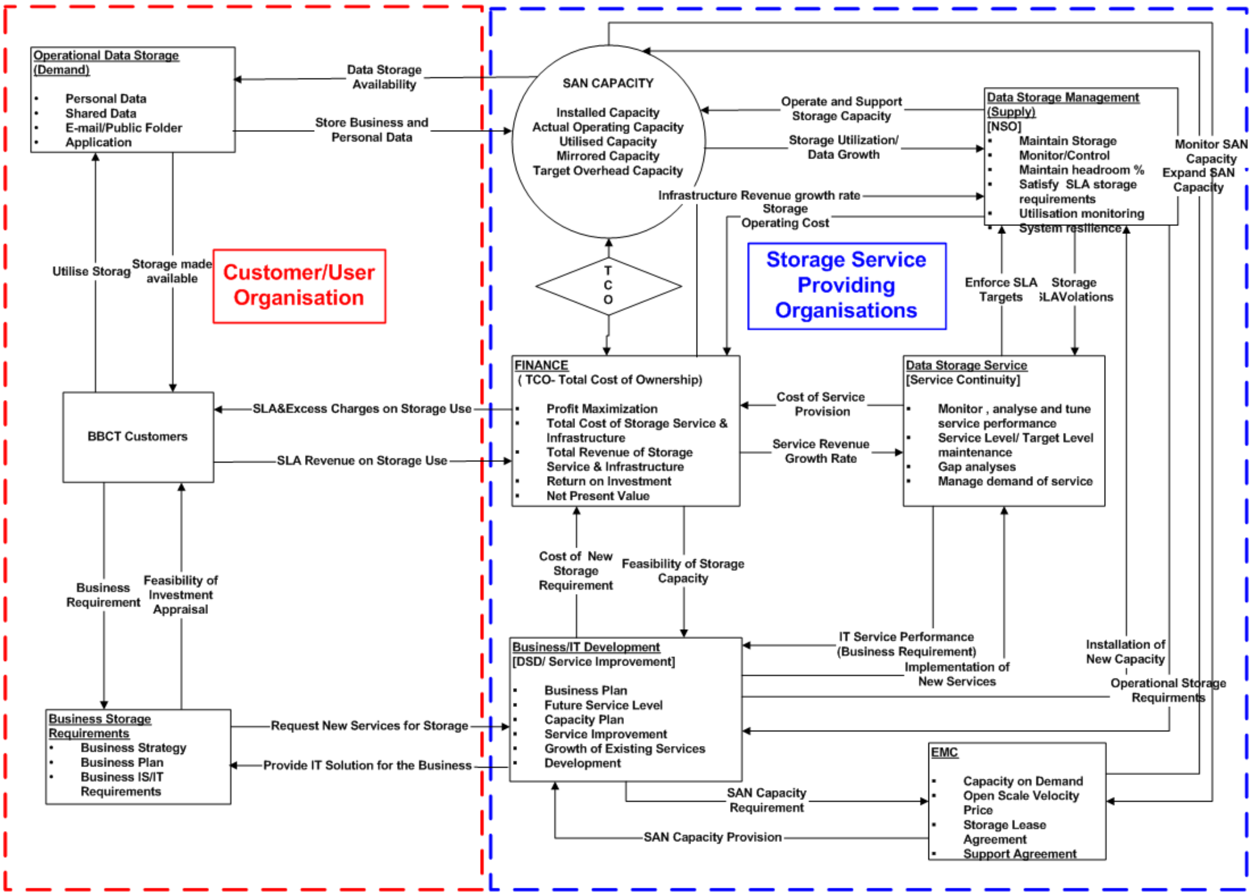
5.1 Design and Modelling Issues

The model design is attempting to address the following key issues: -

1. The understanding of data storage capacity in the context of IT service management and service quality improvement.
2. Introducing simulation as an alternative to current linear projection and decision making methods that are not addressing the exponential demand escalation and applying effective capacity management measures.
3. Using system dynamics to create understanding of the feedbacks in storage demand and understanding the effects of formulating policies, leverages and decisions.
4. Integration of strategic IT infrastructure management with defined IT service management processes.
5. Provide a forum for addressing storage management for a large enterprise that uses data assets in mission critical environments.
6. Defining and investigating key variables in the IT infrastructure management that is critical to efficient and optimised utilisation of data storage capacity.

The model is divided into to conceptual areas of business relationships. These are the Customer/User Organisation and the Storage Service Providing Organisation as illustrated in Figure 5 below.

Figure 5 Conceptual Frameworks of the SAN Capacity Requirement Model



The BBC desktop users are able to generate data storage requirements through two different mechanisms. The first is through the daily operational requirements through the core network services as defined in the model personal, shared, e-mail data storage requirements. The second is through strategic business requirements that are generated by the BBC business strategy, the business plan and over all strategic business IS/IT requirements. The Technology Direction Group (TDG) in the BBC controls this area.

The demand generated by the user/customer organisation affects different stakeholders in the storage service-providing organisation. The model design needs to address the different and at times conflicting interests of the stakeholders. Table 1 below will summarise the various aspirations and interests to different stakeholder in the storage capacity requirements decision-making process.

Table 1 Main Stakeholders in the Capacity Requirements Model

Stakeholder	Aspiration and Interest
Customer/User	Require an IS/IT system that is reliable , effective, fast , easy to use and productive
Data Storage Infrastructure (SAN) Management	Operational manage the SAN system and server infrastructure against a specified SLA with capacity and availability target as key performance indicator want to meet SLA target and optimise storage
Data Storage Service Management	Management of the SLA, monitoring the target are met and analysing exceptions and SAL violations. Need to maintain service quality in consistent fashion.
Business/IT Infrastructure Development	Development of the storage services, capacity acquisitions, server commissioning. Evaluation and introduction of new technology. What to strategically improve systems and services.
Finance Management	What to reduce total cost of ownership , optimise return on investment and maximise revenue and profit
EMC Corporation	What to provide quality service to BBCT, maximise revenue and profit and maintain strategic business partnership with BBC and avoidance of service penalties

The interaction between these competing and at times conflicting sets of interest and aspirations between the various stakeholders form the bases by which model elements are constructed and the process structures and characteristics are defined in the model.

The field study was carried out during the months of March – October 2003. The field study carried attempted to understand the Customer/User organisation and impact of effective capacity management form the perspective of service quality provision. Operational root-cause analysis documents were examined in the field environment to determine persistent capacity related problems stretching over a number of years. Focussed interviews were used to determine the major drivers for the storage service-providing organisation. The study needed to understand the critical success factors and key performance indicators in determining capacity requirements levels and identify the key variables that are used as inputs and outputs to the model.

The model had both hard 'quantitative' and soft 'qualitative' data inputs that were used to create the desired capacity requirement data outputs for the planning of capacity requirement decision-making process. The 'hard' data collection was centred on the customer/user management process, capacity demand generation process, the SAN capacity supply management processes, service delivery control process and the financial appraisal and management processes. The 'soft' factors are of major importance in the capacity requirement decision-making process that are not tangible and therefore are difficult to quantify. In System Dynamics modelling and simulation it is possible to incorporate soft variables with the use of graphical functions by assigning variable values on scale of 0 –1.

6. Implementation of BBC SAN Capacity Requirement Model

System Dynamics (SD) methodology enables the modelling of data storage utilisation and requirements planning processes. Using SD models it is possible to identify the major system variables and their measures of interaction within IT Service Management domain. It is possible to understand the behaviour of users' storage requirements and the consequences of capacity planning decisions using, 'what if?' dynamic analysis to optimise capacity acquisitions for IT service providing organisation. The model could be used to identify optimised capacity levels by balancing the total cost of ownership against capacity acquisitions and supply of storage to users against demand made on the IT storage infrastructure. (Williams 2002) *SD models are rational structures that generate a formal representation of behaviour of the system being studied*. Using SD modelling it will be possible to explore various approaches in providing the required capacity to the business so that the agreed level of storage service is maintained to users which heavily rely on the availability of storage capacity at the right time, at a price acceptable and competitive in the IT Service provision industry. The key question is how the effectiveness of storage capacity requirements planning can be increased to improve the quality and performance of IT service provision to the BBC, while reducing the cost of capacity acquisitions by optimising capacity utilisation and accurate future requirements projection.

The model development is based on the ITIL capacity management process framework. ITIL provides a generic structure in IT service management that can be used to identify the entity classes which enables the construction of model segments and sectors. The fundamental building blocks are the formulation of a relationship between the user/customer organisation and the service providing organisation connected through a capacity demand and supply processes

The model is divided into individual subsystems that are based on the conceptual frame work defined previously in this paper and the key system variables captured in the causal loop diagram formulated during the development process.

6.1 Model Sectors

The model has four interacting subsystems and these are classified as SAN Capacity Demand Management, SAN Capacity Supply Management, SAN Capacity Service Management and SAN Capacity Finance Management. Their associated 17 sectors are presented on Table 2. This represents the problem statement and dynamic hypothesis and discussed earlier. The sector names are based on the ITIL framework that identifies the core processes within capacity management activity. Table 2 summarises the major function of each sub system and their sectors represented in the model.

Table 2 Model Sectors and their Main Function in the SAN Capacity Requirement Model

Subsystem	Sector	Major Function
1. SAN Capacity Demand Management		
	1.1 User Management	The main function is to monitor and quantify the aggregate demand generated on the SAN by BBC desktop users data production.
	1.2 Personal Data Storage Demand	
	1.3 E-mail Data Storage Demand	
	1.4 Shared Data Storage Demand	
	1.5 Data Archiving	
2. SAN Capacity Supply Management		
	2.1 Personal Data Servers Capacity Supply	The main function is to control the supply of storage capacity on the SAN to satisfy demand. Provide storage requirement decisions based on utilisation and availability.
	2.2 E-mail Data Servers Capacity Supply	
	2.3. Shared Data Servers Capacity Supply	
	2.4 Operational SAN Capacity Requirement	
	2.5 Strategic SAN Capacity Requirement	
	2.6 EMC SAN Capacity Reserve	
3. SAN Service Management		
	3.1 Service Level Management	The main function is to maintain the Service Level Agreement Targets on the SAN. Monitor, analyse and tune service performance. Manage the demand of storage services.
	3.2 Storage Service Performance	
4. SAN Capacity Finance Management		
	4.1 SAN Capacity Cost	The main function is to manage the finances associated with capacity utilisation and acquisition. With focus on total cost of ownership, profit maximization and return on investment.
	4.2 SAN Capacity Revenue	
	4.3 Excess Storage Charging	
	4.4 Service Penalties	

The model sectors presented in Table 2 consist of approximately 238 equations in STELLA v8.0 and of this 23 are levels and 51 are rates. There are

17 interacting sectors contained in the four subsystems presented above. The SAN capacity management process environment is divided into four major sub-systems. The SAN Capacity Demand Management subsystem primary objective is to monitor and quantify the load or storage requirements that will be placed on the SAN infrastructure by the BBC user base. The SAN Capacity Supply Management subsystem primary function is to supply capacity required for SAN Capacity Demand Management subsystem by analysing and projecting data growth rate across the server estate attached to the SAN infrastructure. The SAN Service Management subsystem primary function is to control the service delivery quality through key performance indicators that are specified in the Service Level Agreement (SLA). The SAN Capacity Finance Management subsystem is responsible for all finances management activities associated with the utilisation and acquisition of SAN capacity.

6.2 Key Model Equations

The model has seventeen endogenous interacting sectors within the four identified subsystem. Using STELLA software version 8.0, the variables identified are used to characterise the capacity management process and their relationships are explicitly defined and modelled for simulation experiments. The model simulates for 15 years planning horizon with DT set at 0.25 indicating that monitoring activity summary is carried out on a quarterly basis in a year period.

The equations for the variables New_Customer_Request_Rate and Old_Customer_Removal_Rate define the rate at which users accounts are added to the system increasing the demand for storage resources and rate at which users leave the organisation hence releasing storage resources. The interaction of these equations defines the level of BBC Customers that will generate storage capacity requirements on the SAN.

New_Customer_Request_Rate = Normal(Average_New_Customer_Request_Rate,BBCT_Customers)/Customer_Request__Adjustment_Time {user/yr}

Old_Customer_Removal_Rate= Average_Account_Deletions_per_month/Account_Deletion__Adjustment_Time {user/yr}

The type of data stored on the SAN can only be generated from three file server types, which connect to the SAN forming the core network business services for the BBC. These are personal data (FS) servers, e-mail (XU) servers, and shared data (RD) servers. The three server types are incorporated in the model to fully analyse storage demand put on the SAN. The demand generated is characterised as data production and defined in the equations below and these key equations determine the level of data storage requirement in the demand management subsystem. The storage demand generated is satisfied through the installed server infrastructure using the

SAN capacity supply management. The equations below represent the total data production rate for the three server types that connect on the SAN.

Total_FS_Data__Production= Average_No__of_Customers__per_FS_server*Number_of__FS_Servers*Data_Generated__per_customer {Megabyte}

Total_XU_Data__Production= Average_No__of_Mailbox_per_XU_server*Number_of__Exchange__Servers*Mailbox_size_per_user {Megabyte}

Total_RD__Data_Prodution = Average_No__Customers_per_RD_server*Number_of__RD_Server*RD_usage__per_Customer {megabyte}

The three equations below characterise the data growth rate factors that are observed in the field on the SAN infrastructure and are used to adjust the required level of storage supply for the three server types connected to the SAN .

FS_Data__Growth_Rate= CGROWTH(FS_Data_Growth__Rate_Factor)*FS_Server_Data {megabyte}
XU_Data_Growth_Rate = CGROWTH(XU_Data_Growth__Rate_Factor)*XU_Server_Data {megabyte}
RD_Data_Growth_Rate = CGROWTH(RD_Data_Growth__Rate_Factor)*RD_Server_Data {Megabyte}

These growth factors for the different server types determine the level of storage requirement generated by interacting with an example equation on the personal data server defined below. The servers are operationally monitored for SLA targets and operational requirements are generated based on the availability thresholds defined in the SLA. This is calculated using the following example equation on the personal data FS servers

FS_Data_Storage__Requierment = Actual__FS_Data_Storage__Availablity-Target_FS__Data_Storage_Availablity {megabyte}

The capacity supply sectors operates using key understandings in operational storage management dynamics that differentiates total capacity increase through the expansion of existing servers or through the installation of new servers. Both have the effect of increasing total usable storage capacity but through different supply lines. These dynamics are captured in the following equation.

Expanding_XU_Servers_Capacity = (Total_XU__Servers_Capacity+XU_Server_Capacity_Expansion_Requierment)/XU_Server_Capacity_Adjust

Exchange_Servers_XU_Allocating = (XU_Server__Mirroring+Storage__Requirement__per_XU_Server)*Number_of__XU_Servers/Exchange_Server_XU_Allocation__Time

There are differences between the processes of expanding the capacity on an existing operational file server and allocating or installing new file server in the SAN. When installing a new server additional capacity is needed for mirroring, resilience and replication which are captured in the example equations used for the E-mail Data Servers Allocating variable. The above equations are used to supply all storage requirements generated from the demand management sectors.

Operating SAN capacity stock is defined to provide storage capacity for all the three server types requiring expansion due to the service level violations resulting from high utilisation thresholds. This is expressed in the equation below.

Operating_SAN_Capacity(t) = Operating_SAN_Capacity(t - dt) + (Supplying_Operating__SAN_Capacity - Expanding_FS_Servers_Capacity - Expanding_RD_Servers_Capacity - Expanding_XU__Servers_Capacity) {Megabyte}

The available capacity stock defined in the equations below is the source of capacity supply to new projects and allocation of new server capacity for all the server types defined in the SAN. The two equations for Operating SAN Capacity and Available SAN Capacity are at the centre of the SAN capacity requirement model representing the relationship between the two most important system variables of the model. These two variables are the key decision making variables and together provide the model with the total SAN capacity stock that can be managed by BBC Technology. These two variables operate together to satisfy the storage demand requirements of the BBC.

Available_SAN__Capacity(t) = Available_SAN__Capacity(t - dt) + (Expanding__SAN_Capacity - Supplying__New_Projects_Capacity__Ca - Resource__Data_Servers_RD_Allocating - Personal__Data_Servers_FS_Allocating - Supplying_Operating__SAN_Capacity - Exchange__Servers_XU__Allocating) * dt {megabyte}

The equation below Capacity requirements on SLA violations is a level that represents the total utilisation of capacity above the SLA threshold. This stock is an input to the rate of operating capacity supply

Capacity_Requirement__on_SLA__Violations(t)=Capacity_Requirement__on_SLA__Violations (t - dt) + (Rate_of_Capacity_Requirement_on_SLA__Violation) * dt
INIT Capacity_Requirement__on_SLA__Violations = 0

The revenue generated in the model is expressed as a level using a rate equation defined below. The rate is determined by the variables defined in the equation Total Excess Storage Charges, SLA Revenue from SAN Customers, SLA Revenue from the infrastructure and revenue generated from project activities. The revenue from SAN customers are dependent upon the number of users singed up on the Desktop SLA and SLA Storage Charges per customer which are parameters observed in field study.

SAN_Revenue Rate = (Total_Excess__Storage_Charges+SLA__Revenue_SAN_Customers+StorageInfrastructure_SLA_Revenue+Storage_Revenue_on__New_Projects)/Revenue__Adjustment_Time {£/yr}

The equations defined below defines the Total Cost of Ownership of the SAN infrastructure. Together with the storage lease cost defined in the model determine the level of total SAN capacity cost.

SAN_Operating_Costs = Human_Resource_Cost+SAN_Accomodation_Cost+SAN_Hardware__Maintenance_Cost+SAN_Software__Cost+SAN__Power_Consumption_Cost {£}

6.3 Simulation Results

The simulation results can be directly compared with the reference mode behaviours identified in Figures 2 and 3. The level of operating SAN capacity increases as the available SAN capacity decreases. The data growth rate is expected to rise as the number of BBC users increase and this will increase the level of capacity requirements.

Simulation results were generated using ten key variables that have been defined in the dynamic hypothesis. Figure 7 and 8 show the simulation outputs using the model. Figure 7 overleaf illustrates the relationship of Operating SAN Capacity (1), Data Growth Rate [represented by FS Server Data (2), XU Server Data (3), RD Server Data (4)] ,BBC Customers (5). The output from the simulation reproduces the base case behaviours in the dynamic hypothesis.

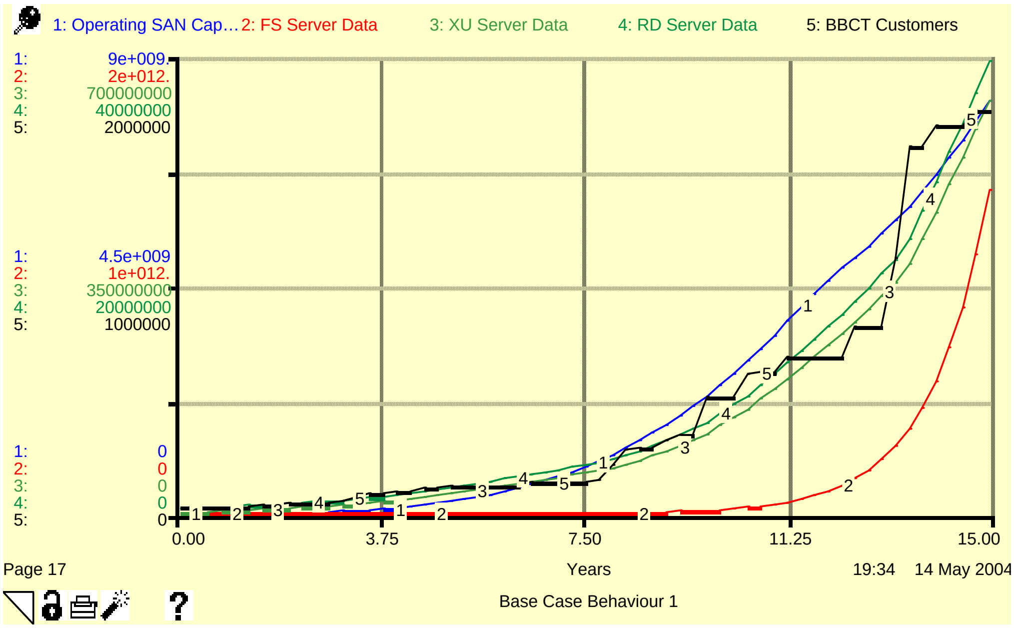


Figure 7: Simulation Outputs BBC Customer, Data Growth Rate and Operating Capacity

Figure 8 presents the simulation output that shows the behaviour of key variable Capacity Requirement on SLA Violations (1), Available SAN Capacity (2), Total FS Server Capacity (3), and Total Exchange Capacity (4) and Capacity for New Projects (5). It is possible to see a behaviour defined by one of the key propositions in the dynamic hypothesis that when Capacity Requirements (1) increases Available SAN Capacity (2) will being to decrease

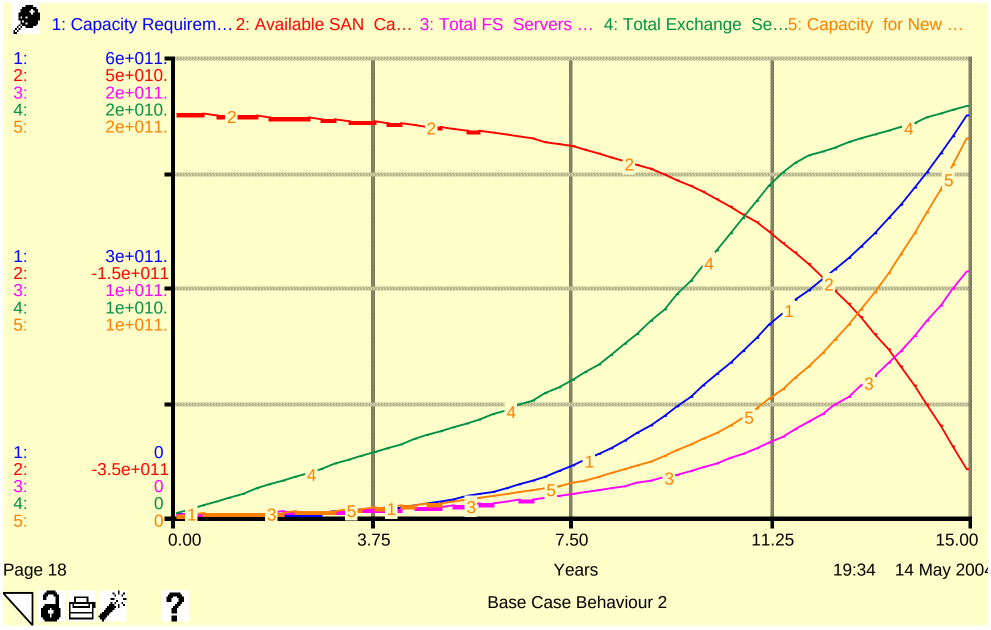


Figure 8: Simulation Outputs Capacity Requirements, Available Capacity and Total SAN Capacity

6.4 BBC Case Analysis and Model Credibility

The case analysis attempts to show the SAN Capacity Requirements Model is capable of generating plausible data that is consistent with the reference modes and dynamic hypothesis defined in this paper. The analysis correlates data that is actually generated in the field of operation at the BBC with data generated through model simulation. It assesses the model behaviour against the operational environment to gain insight of the accuracy and credibility of the implemented model.

The data available in BBC Technology that was recording utilisation trends and data growth rate behaviours was available on Personal Data (FS) servers. The data is automatically gathered from all FS servers in London where seven FS servers are on the SAN infrastructure. The FS servers installed in the SAN have a total capacity of 150 Gigabyte (150,000 Megabytes) and the servers outside the SAN infrastructure (i.e. stand alone servers) are installed with 200GB or 100GB. On a fortnightly basis the servers' utilisation is measured against total server capacity and the figure is converted into % utilisation against the total server capacity. These figures are recorded in Excel and a trend analysis is carried out using a Time Series Analysis. The diagram on Figure 9 presents the actual time series analysis done on sixteen FS servers in London showing the utilisation over time behaviour. The x-axis presents the time in months and y-axis presents the % utilisation.

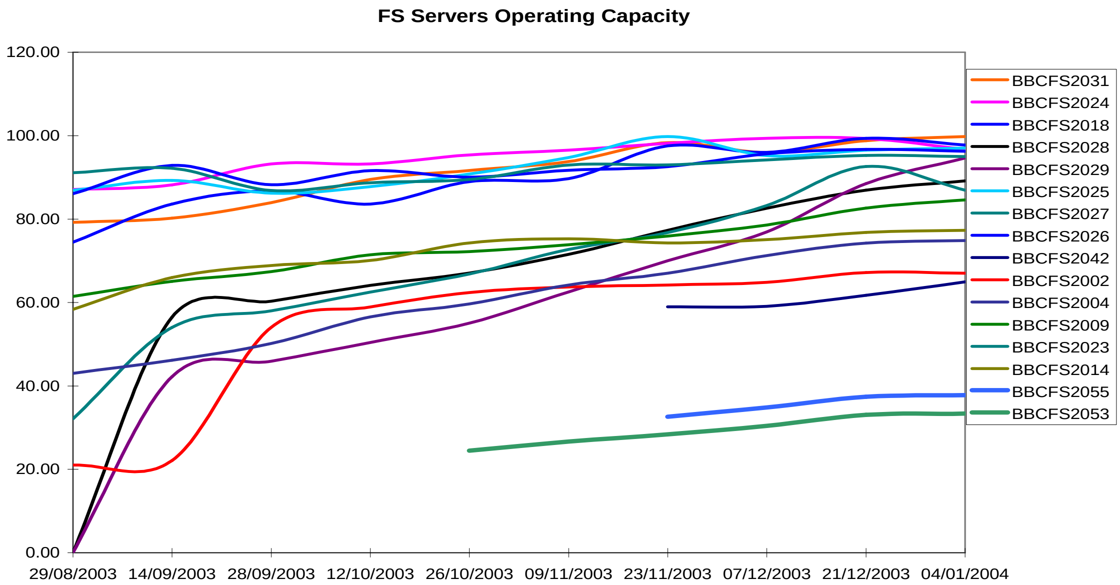


Figure 9 Actual FS Server Operating Capacity at the BBC

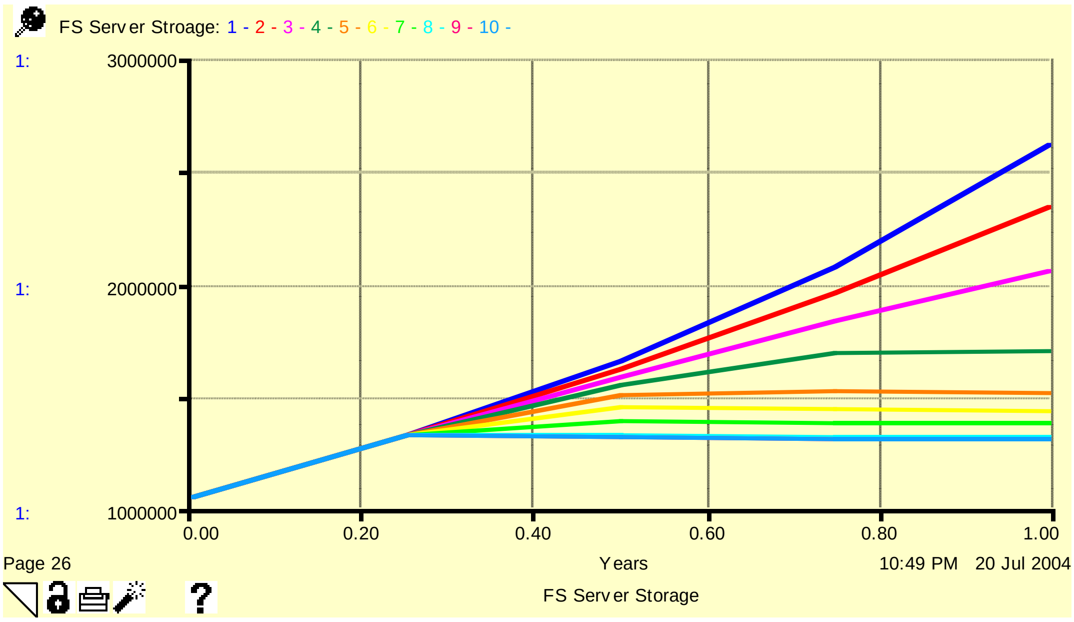


Figure 10: Simulation output for FS Servers Capacity and Storage Utilisation

Figure 10 presents comparative simulation outputs of the SAN Capacity Requirements Model on FS Server Storage over a one year period.

The graph on Figure 9 shows the behaviour of live operational servers using data gathered at the BBC and converted in a time series analysis with linear approximation for a period a year. The graph on Figure 10 captures the behaviour of seven SAN attached FS servers using the system dynamics simulation model over a period of a year. It is possible to see the similarity in behaviour as both graphs show an increase in capacity utilisation with exponential characteristics. The similarity of the simulation result and the actual case analysis indicates that SAN Capacity Requirements model is able to capture real world business systems behaviours observed by the model through the dynamic hypothesis. This finding confirms the suggestions on the usefulness of system dynamics in capturing the intricate dynamics based on feedback theory of the capacity requirements process.

7. Conclusions

The research described in this paper models the SAN capacity requirements planning and management dynamics from a feedback perspective. The success of balancing the demand and supply of data storage capacity at the right price, while optimising the utilisation of installed operational capacity is critical to the competitiveness of the IT services providing organisation. Its competitiveness can increase by reducing the total cost of IT storage service provision and increasing the level of service quality it provides to its customers. The efficient and effective delivery of data storage capacity is crucial to the performance of organisations that strategically depend on the supply and availability of data storage capacity. The proliferation of information systems and information technology as the backbones of business operation make data storage capacity management a critical success factor in strategic business management.

In response to these critical business requirements, IT service providers have deployed high cost infrastructure monitoring tools to capture the utilisation trends of the data storage systems and provide data analysis and tuning recommendations in the operational environment. However the tools have not been able to fully understand or capture the underlying business data storage requirements or data production capabilities of the customer organisation. Therefore the tools are only able to provide reactive capacity planning functionalities based on monitored historical data using linear projection techniques. The tools are not able to account for the feedbacks from business requirements and associated demand side variables. The linear demand projection methods have shown severe weaknesses applied on exponentially volatile behaviours in data storage requirements. In this paper the SAN Capacity Requirements Model applies the system dynamics approach using the Dynamic Synthesis Methodology. It captures the feedbacks generated with changing business requirements and associated data production rates to formulate a clearer view of strategic storage capacity requirements.

The research found that the level of data production combined with the number of users accessing the storage infrastructure, the type and number of desktop applications accessed by users highly impacted the levels of operating capacity and available capacity. This indicates that analysing just the level of data growth rate is not enough to make an accurate analysis of capacity requirements. The experiments conducted on the system dynamics model shows the level of operating SAN capacity is not sufficient to quantify the level of business requirements for data storage. The level of operating SAN capacity is dependent upon the finite level of available SAN capacity, the sum of these two variable represent the total SAN capacity installed on the storage infrastructure. This cause and effect relationship between Operating and Available SAN capacity often excludes the users underlying root cause dynamics of business requirements and data production capabilities resulting in flawed data growth projections. The most likely outcome is inadequate or excessive capacity acquisitions. The model is able to generate various outputs for data growth rates and levels of SAN capacity based on storage demand and data production rates, which are the key variables in data storage requirement analysis. This is the first time this link has been analysed, quantified and simulated in the organisation.

The understanding of IT capacity management and the importance of business requirements analysis has been highlighted in structured IT service management approaches. The ITIL framework defines modelling as one of the key capacity management activities. This paper reinforces the application of system dynamics modelling as a basis of strategic capacity requirements decision making by understanding the feedback loops present in the capacity requirements process. It is this feedback loops that research and practicing managers can exploit by aligning operational IT knowledge, technological innovation and organisational processes to reduce requirement uncertainty and effectively optimise capacity acquisition and utilisation.

The SAN Capacity Requirements model could be the basis for ITIL based capacity management modelling approach to provide a holistic understanding of the business data storage requirements dynamics. The model can be used to investigate the effects of volatile and unpredictable changes in data storage requirements resulting from rapid changes in information systems development and technological innovation.

References

Boehm, B., (1981), Software Engineering Economics, Prentice Hall, Englewood, Cliff NJ

Burd, S.D. and Kassicieh S.K. (1991). *'Decision Support for Supercomputer Acquisition'* , Journal of Operations Research, Vol. 39, No 3, May-June

Checkland, P., and Scholes, J. (1990). " Software System Methodology in Action". Wiley, Chichester.

Daellenbach, Hans G.,(1999). Systems and Decision Making – A Management Science Approach, John Wiley & Sons

Dataquest 2000, Gartner Business Information, www.gartner.com access at 20.09.03

Davis, C. and Vick, C R. (1997), *'The software Development System'*. IEEE Trans Software Engineering

EMC Corporation, (2002), BBCT SAN Final ITT Commercial Response

Kleijnen, J. P.C. (1980). Computers and Profit: Quantifying Financial Benefits of Information. Addison Wesley, Reading Mass., 28-29

Ku A. (1995). Modelling Uncertainty in Electricity Capacity Planning – Phd. Thesis for London Business School, University of London.

Leman Brothers. (2001), 'Enterprise *Storage Takes Centre Stage, What will be the Encore?'* Leman Brothers.

Maani Kambiz & Casvana, R. Y., (2000). Systems Thinking and Modelling, Prentice-Hall New Jersey

Meadows, D. M. (1982). Lessons from Global Modelling and Modellers Futures Vol:14 pp:111-121

Microsoft Excel Corporation, Microsoft Excel 2002 Help, www.micrsoft.com accessed at 20.06.04

Pidd, M (1992). Computer Simulation in Management Science. John Wiley and Sons Ltd,UK.

Richmond, B. (2001), An Introduction to System Thinking, STELLA, High Performances Systems Inc, Lebanon, NH.

Senge, P. M. (1990), The Fifth Discipline: The Art and Practice of the Learning Organisation. Doubleday, New York.

Sherwood, D. (2002). Seeing the Forest for the Trees- A Manager's Guide to Applying Systems Thinking, Nicholas Brealey Publishing, London.

Stevens, B. A. 1980. Audit and Control of Performance in Data Processing. EDP Perform. Rev. 8, 1-13

Stewart,J. (1990). Capacity Planning, Central Computer and Telecommunications Agency (CCTA), John Wiley & Sons, Inc, New York

Williams, D. (2001). A System Dynamics Theory of Requirements Engineering Process. In proceedings of the International System Dynamics Conference 2001, Atlanta, USA. 23 –30July 2001

Williams, D. (2002) '*Dynamic Synthesis Methodology: A theoretical framework for decision process modelling and* analysis'. In the proceedings of the 1st European Conference on Research Methods for Business and Management Studies, (Ed) Dan Remenyi, MCIL, Reading 29-30.

Williams, D., (2003*)* ' *Challenges of System Dynamics to Deliver Requirements Engineering Projects: Faster, Better and Cheaper*'. In proceedings of the 21st International System Dynamics Conference, New York, USA, 20-24 July 2003, System Dynamics Society

Williams, D., Hall, T., Kennedy, M., (1999). '*A Framework for Improving the Requirement Engineering Process modelling*'. In Eds C. Hawkins, G. King, M. Ross, G. Staples, Managing Quality: Proc of Software Quality Management Conference Southampton UK 29-31 March.

Wolstenholme E. F. (1990) "System Inquiry – A System Dynamic Approach", Wiley, Chichester