IST-Africa 2008 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2008
ISBN: 978-1-905824-07-6

Evaluating Health Information Systems for Developing Countries using Simulation

Kieran SHARPEY-SCHAFER, Hussein SULEMAN
Department of Computer Science, University of Cape Town, Private Bag X3, Rondebosch, Cape Town 7701, South Africa
Tel: +27 21 650 2663, Email: ksharpey@cs.uct.ac.za

Abstract: Digitization and networked computing in the healthcare sector have resulted in electronic patient records that are stored, managed and shared among different healthcare providers. In Developing Countries such systems are being considered to improve on healthcare service delivery, with the aim of nationally available patient records. To implement this, network architectures and data transfer solutions can be adapted from other contexts, such as centralised or peer-to-peer computing. However, it is not always clear that such solutions are most appropriate, especially given the unstable and limited resources of developing countries. This paper proposes that discrete event simulation can be used as the foundation of a tool to measure and evaluate the performance of algorithms and network architectures. The initial proof-of-concept tool is presented, along with how it can be used to evaluate a solution for sharing records between a group of facilities. Initial results are promising and further envisaged investigation is discussed.

Keywords: Electronic Patient Records, Developing Countries, Simulation, Evaluation of Networks.

1. Introduction

Electronic Patient Records (EPRs) are increasingly being investigated to improve healthcare delivery, by storing and tracking medical data over the lifetime of a patient, typically across multiple healthcare centres. EPRs can arguably increase quality of care through improved sharing of information, broader standardization, quicker data retrieval and automated analysis support [1].

A nationwide implementation of patient records, sometimes called a National Patient Record (NPR), would allow the monitoring of healthcare delivery across the country, disease monitoring and nationwide access to patient histories. Many developing countries, including South Africa, have begun investigating possibilities for NPR systems, with varying degrees of success. This comes at a time when many so-called developed nations are at the implementation or roll-out stages of national systems.

Numerous failures of both EPR and NPR projects indicate that the planning and analysis of these solutions is increasingly important. This is particularly true in developing nations where IT infrastructure is still under-developed, especially in comparison to the countries that the IT solutions often originate from.

Common problems include unstable power in rural areas, unreliable data transmission over networks, failures for extended periods, uneven bandwidth availability on different network segments and often a complete lack of sufficient bandwidth for many primary healthcare facilities. While developing countries often look to their northern neighbours for ICT solutions, these localised infrastructural problems can introduce drastic operational or performance challenges into the system. This is an example of a 'design-reality' gap, which is when factors of the real world implementation differ from those considered in the design, leading to operational difficulties.

The subject of this paper is to introduce a technique to evaluate whether or not specific data management solutions will be effective and efficient in developing country environments. The approach suggested is one of discrete event simulation – by simulating the nodes and arcs of a networked distributed or centralised healthcare information system, it is possible to measure and study the interaction effects of different algorithms and base network conditions. Algorithms that emanate from, for example, peer-to-peer networking or hierarchical metadata harvesting may be applied to different local network infrastructures, thereby making it possible to assess the suitability of these widely-promoted approaches in a realistic developing country scenario.

This approach is restricted in its evaluations to quantitative techniques mainly focused on the technical performance of the EPR – which have some direct influence on various stakeholder interests. Some of these could include: performance, accuracy or reliability. There are obviously many other important factors that influence the 'success' of an EPR; usability, maintainability, organisational acceptance, etc. However these are beyond the scope of discrete event simulation techniques and could perhaps be evaluated separately. Also not considered in this study are the practical concerns of parallel paper and electronic systems, skills development and other such human factors, while acknowledging that they are crucial to systems in practice.

It is hoped that the methodology illustrated here will allow proposed EPR solutions to be evaluated more accurately, using the quantitative methods shown, to establish the suitability in a particular developing context. Furthermore the simulations could possibly elicit operational barriers that were difficult to foresee in the original design. A real-world application of its use could include high-level strategic health professionals considering funding new solutions could request simulation of the solutions. Architects of HIS could also use the simulation to assure proof-of-concept when proposing a new solution or even an existing solution in a new context. Furthermore a system could be evaluated under 'non-optimal' conditions such as a natural disaster, disease outbreak or the growth of populations or epidemics.

2. Background

2.1 Benefits of EPRs and NPRs

A study of South African patients performed by Accenture in 2006 [2] showed that almost a third of all participants used multiple healthcare providers and half said that they answered the same questions on each visit to a new practitioner. Access to a reliable NPR would make medical histories quicker to obtain and more comprehensive, giving healthcare workers more time and information with which to do their job. This could reduce the number of questions, and thus the visitation times, as well as potentially improve care as a patient history completed by another health professional would presumably be more accurate than that recounted by a patient.

Government also can benefit from the implementation of an NPR, by having access to wider, consistent, up-to-date aggregate data. Without a national network providing access to health information, any data that is captured, whether in electronic or paper format, only resides at the healthcare facility in which it is used. Theses silos of clinical information are difficult to survey and thus make it difficult for the government to monitor healthcare delivery as well as potential disease management [3].

2.2 NPR Solutions

There are many possible technical solutions for creating an NPR. Both existing and proposed NPR designs are taken from many different IT fields including: Web applications, distributed databases, peer-to-peer networks, digital archives, interoperability standards, and grid computing.

The disparity in approaches is evident in some of the world's largest proposed systems, in the USA, UK and Brazil. The USA has chosen to implement a decentralised model [4], where facilities have access to a regional gateway which in turn provides access to other regions (Figure 1). A patient's EPR is stored at a local facility and can be requested by remote parties. Alternatively, the UK has a centralised solution that allows all facilities nationwide to access the central hub of patient records. Lastly, Brazil uses a combined approach – a hierarchical network that has facilities passing patient data up to regional centres that then replicate data to a national server.

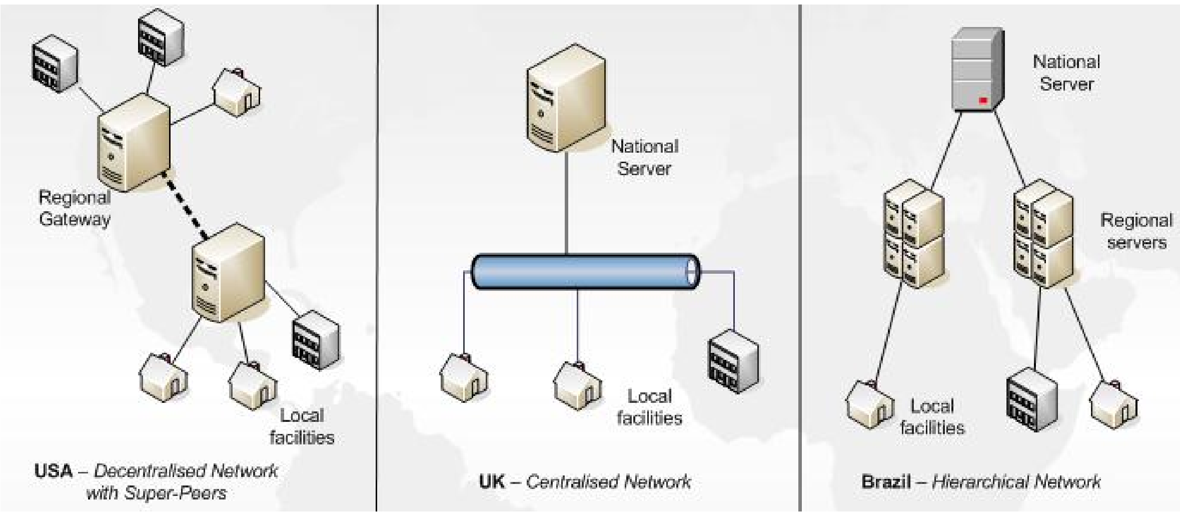

Figure 1: Examples of the Different International NPR Designs

A centralised solution's main benefit is simplicity and consistency of data management [5]. As data is in one location, and presumably in one format, this allows quicker analysis, consistent data and easier access control. However there is some risk in implementing a centralised solution as the entire network will be affected by the central node's reliability. Furthermore a centralised solution can overlook local characteristics and differences in the capacities of local facilities.

A decentralised network has no central repository and allows facilities to retain records and share these on request. Decentralised networks can be purely *peer-to-peer* where all parties in the network are equal and follow the same protocol for interaction with all nodes. Alternatively, many decentralised networks have super-peers, which are nodes that are given some responsibility to co-ordinate communication in the network – often as simple as listing which nodes are in the network. A good example of this is the proposed NPR solution in the USA where facilities have access to a regional gateway which in turn provides access to other regions. The primary benefit of a decentralised solution is that it removes the reliance on a single node. If parts of a peer-to-peer network fail, the other nodes can continue sharing data. However, co-ordinating consistent records across a decentralised system can require much effort, and collecting consistent aggregate data from the system is more complex than in a centralised design.

2.3 Challenges of the Developing Country Context

Technical infrastructure in many developing countries often lags far behind others in terms of capacity, coverage and consistency. Examples of this include the intermittent phone lines and power at many facilities, the lack of IT resources in many facilities, and the inconsistencies in technology used among various healthcare providers. Furthermore the difficulties in national healthcare in many DCs can also affect the success of an IT solution. Shortages in skilled staff reduce the capacity of many facilities, and the effects of disease and poverty often can create different patterns of healthcare usage among the population.

Taking these contextual differences into consideration, it can be reasoned that solutions that succeeded elsewhere may have different performance issues in a DC. In many efforts the importance of response times, reliability, security and data integrity of the system has been stressed [4] [6].

3. Methodology

In the following section the methodology used to evaluate different NPR solutions in a developing context will be presented. The methodology can be broken down into two sections: modelling of solutions and context; and the simulation of NPR solutions for evaluation.

3.1 Modelling

There were two main aspects of the NPR to model: firstly, the IT architecture and algorithm for providing nationwide access to EPRs; and secondly the load and constraints on the NPR. The model was used to describe the network at a high level using Queuing Network Model (QNM) theory [7] and principally modelled four entities: clients, servers, populations and network connections. *Clients* are nodes in the network representing facilities that are using EPRs for direct care, such as hospitals or clinics. *Servers* are nodes in the network that may provide nationwide or local access to client nodes. *Populations* are used to model the potential population coverage of a facility or region. *Network connections* model the various connections among nodes in the network. Each of these entities is parameterised as follows:

Table 1: Entities of Network Model and Their Parameters

Entity	Parameters
Client	*Number of service points, service time, downtime patterns, population, network connection*
Server	*Dependant client, throughput capacity, downtime patterns, network connection*
Population	*Population, healthcare usage patterns*
Network Connection	*Propagation delay, loss probability, bandwidth*

3.2 Real World Parameters

The attributes for the population model are fairly easy to acquire – many facilities keep records of what geographical area and associated population they cover. Furthermore almost all facilities take their capacity in consideration when planning the allocation of resources. They know, and often report, the average number of patients seen on a day as well as the times of day when the patients tend to arrive. This helps to create the parameter ‘healthcare usage pattern’ which often will be described by a statistical function that emulates the arrival pattern of patients at the facility. These patterns will differ depending on the nature of the facility: a doctor’s office will have a fairly regular arrival rate during office hours, while a public clinic often experiences a major influx of arrivals in the morning. An accident and emergency room at a hospital may experience an altogether more irregular arrival pattern during the course of a day.

The more technical parameters, like bandwidth and loss probability in the network connection, can be generally associated with the standard performance of the relevant technology or hardware. For example, a network connection that uses a dial-up modem cannot expect bandwidth greater than 52kbps and is more likely to experience something lower. Similarly a dial-up connection over a fixed line would experience less loss than connections over wireless solutions such as WiFi, Wimax or GPRS.

3.3 Simulation

Simulation was chosen as the method of analysis as it can provide measurable metrics for assessing the performance of a network and OMNet++ [8] was chosen as the network simulation tool to be used. OMNet++ is an open-source discrete event simulator. A network is created by first defining the behaviour of the nodes in the network and then connecting the nodes together using defined connections. Events at a node – this could be patient or message arrival – stimulate action in a node which can in turn issue more events to the network. Data for evaluation is captured during the simulation and can be monitored during the simulation run or as a summary afterward. The metrics created to evaluate the performance of an NPR solution will be described in the following section.

3.4 Metrics and Evaluation

From the literature on network performance metrics [9], the metrics of greatest relevance for network performance can be grouped in four categories: availability; loss and error; delay; and bandwidth. Availability metrics refer to the percentage of time that resources in the network remain available. This often is expressed as a percentage of time that a particular node or service is expected to be available. Loss and error metrics describe how often messages in the network are lost or corrupted. Delay metrics illustrate the real-time delay experienced in successfully transmitting a message or request in the network. Lastly, bandwidth metrics monitor the data volumes that are transmitted in the network.

Data freshness is another group of metrics that can be considered in a distributed records network that uses caching or other non-real-time access methods to the NPR. Data freshness metrics measure the ‘quality’ of the data being used at a node by analysing how up-to-date it is compared to its copies throughout the system [10].

4. Illustrative Example: Centralised Primary Healthcare Network

This example aims to analyse the performance of a centralised network across a small generic cluster of Primary Healthcare facilities as is typically found in South Africa. The example solution involves an architecture that has a central server storing all EPRs for the health district. Facilities in the area can request a patient’s EPR given that some standard identification process is used. This is not an unusual solution and fits the model of many Web-based EPR efforts where a host allows access to multiple facilities, keeping a single consistent record per patient across the health district.

When implementing such a system, the local health and implementation authorities will no doubt ask: *“What infrastructure will I need to implement the service?”* or more specifically, *“What kind of bandwidth will be needed at the various nodes for this solution”*

To answer this question, a simulation of the proposed network was run under the conditions of an average day’s work for the health facilities, and the bandwidth usage at each of the facilities and the main server was monitored.

4.1 Model

A model was created in OMNet++, as shown in Figure 2, for the proposed centralised architecture. Each node implements its own algorithm to define its behaviour with the help of external events as stimuli. The patient_request nodes' primary behaviour is to simulate the realistic arrival of patients at a facility based on the number of patients expected in a day and when most patients. Using these variables the patient_request node can generate patients 'arriving' throughout the course of the day.

The facility's behaviour then is centred on marshalling patient arrivals into EPR requests to the central server. The central server in return is mainly concerned with replying to each request with the appropriate EPR.

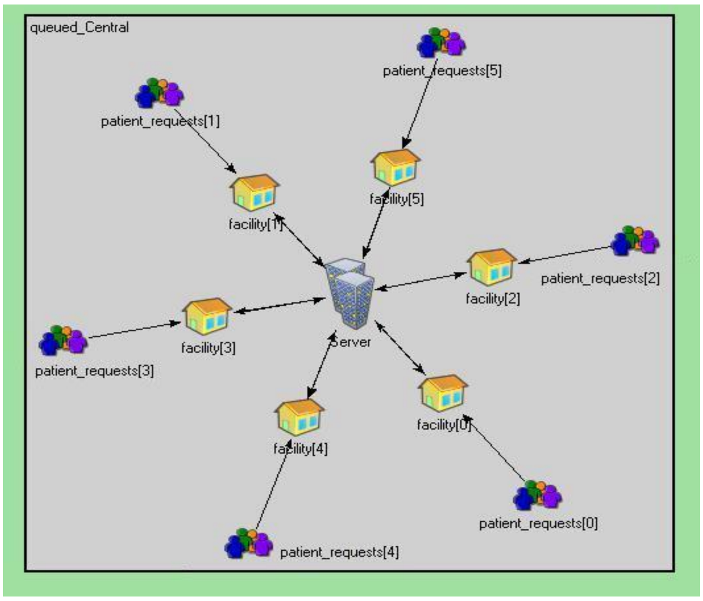

Figure 2 – Centralized Network Supporting 6 Primary Healthcare Facilities

Some real world assumptions where made:

- Simulation duration = 1 day = 8 hours clinic is open
- All facilities serve approximately 80 people per day, who arrive mostly in the morning and tail off toward the later hours (a gamma probability distribution).

4.2 Results

To calculate bandwidth, the number of packets (at an assumed average of 1KB per packet) per unit of time on a node's incoming and outgoing connections was counted (Figure 3).

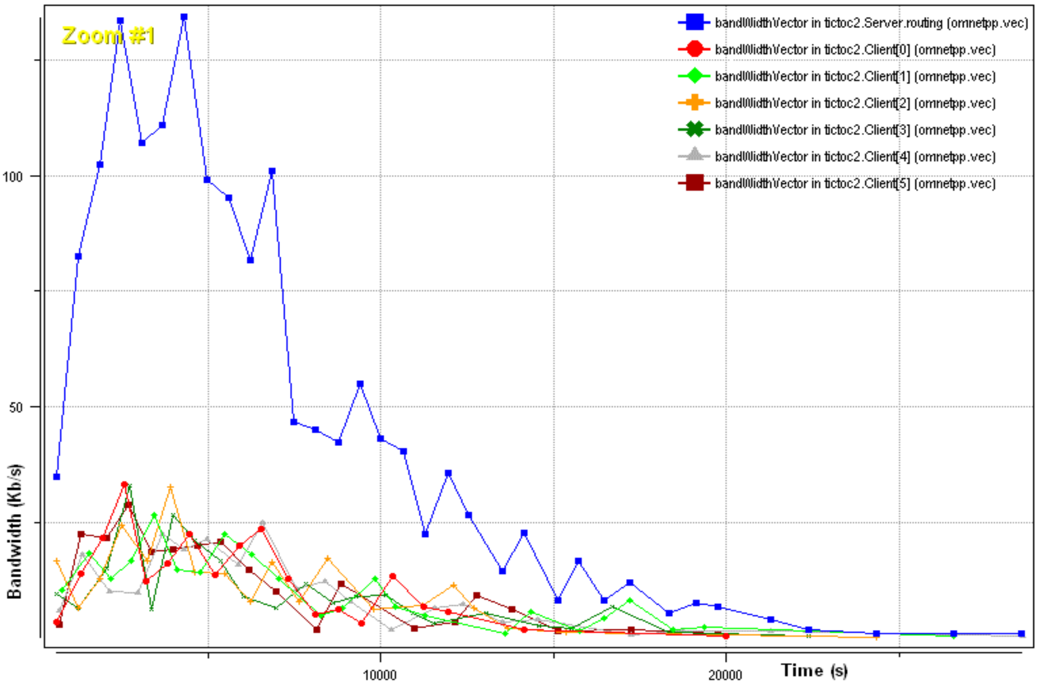

Figure 3 - Bandwidth for 6 Facilities and Server – Bandwidth (Kb/s) vs Time(s)

On analysis it can be observed that the bandwidth follows the same pattern as the arrival distribution, namely some form of a gamma distribution. This is because each request is sent *as a patient arrives at the facility*. This is not always realistic as a normal clinic often has a waiting line for doctor's consultation – where presumably the EPR would be used.

The model was then changed to include service points – modelling the Doctor as a service point of patient requests. This enriches the model as many facilities (e.g., hospitals) will have more than one service point and queues of patients. Figure 4 shows the bandwidth needed for this updated model.

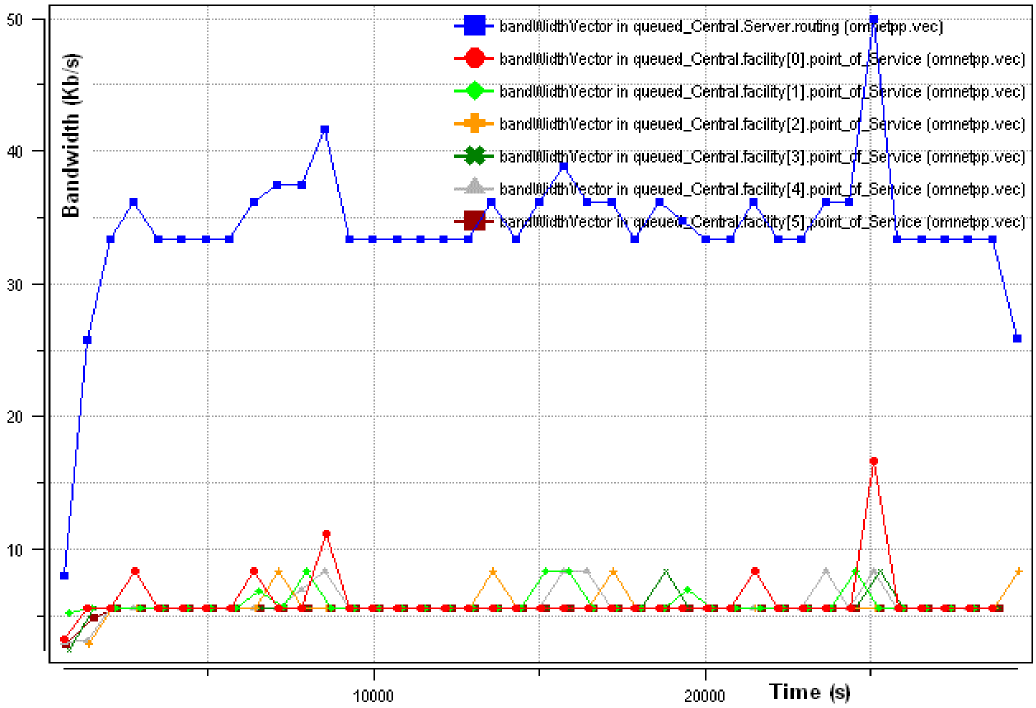

Figure 4 - Bandwidth for 6 Facilities (with Queues) and Server – Bandwidth (Kb/s) vs Time(s)

It now appears as though the bandwidth is reduced to a somewhat uniform distribution. This is plausible as the arrival rate may increase beyond the rate the doctor can process patients. It has caused the bounds of the bandwidth used to drop to depend on how frequently the doctors can process their patient queues.

4.3 Analysis

From analysing the two runs it can be deduced that the internal process of how or when the requests are issued can affect the bandwidth. The second scenario, where the requests are only generated while the Doctor is consulting the patient, is presumably more realistic. However, the process in many clinics currently involves identifying a patient and retrieving a patient's folder *on arrival*. There are many practical reasons for this – it allows the receptionist to keep track of who is waiting and do initial triage if necessary, and the commonly reported phenomenon that many Doctors refuse to enter data into a PC. Thus the initial run (request generation at patient arrival time), is not so outlandish.

Using the model to evaluate the proposed simplistic system has not only produced bandwidth ranges that the system could be expected to require during an average day, but also highlighted how a process detail can create quite a varying demand on the system.

Furthermore, by looking at the quantities of the bandwidth usage at the server we can see that even at the worse case the maximum bandwidth required is in the order of 150KB/s, while the facilities are always below 50kb/s. This would be fairly easy to handle for most modern servers – but makes the distinction that whilst the facilities could potentially survive on dial-up (max 52kb/s) the server would require higher bandwidth, presumably a DSL line.

The use of the simulation tool has thus helped in evaluating the hypothetical healthcare system in terms of scarce resources of developing countries.

5. Conclusions and Future Work

In this paper, it was argued that evaluation of Healthcare information systems can be conducted using discrete event simulation. The method of modelling different EPR solutions has been presented along with some initial results from early simulations of a modelled network. The prototype simulation system is able to measure the service availability, bandwidth usage and response time of nodes in a network, and as development continues further metrics, such as data freshness, will be added.

Future work also includes experiments with more network architectures as well as the scalability of various solutions – can a proposed solution maintain its performance when applied at different regional levels? Further, caching and data harvesting are being incorporated into the different nodes to enable the modelling of a greater range of algorithms for hybrid on-/offline data transfer.

It is hoped that this simulation and modelling approach will ultimately make it possible to design and evaluate NPR solutions for applicability in developing countries before the expensive implementation and rollout phases.

Acknowledgements

This project was made possible by funding from the University of Cape Town, the South African National Research Foundation (Grant number: GUN2073203), THRIP/Telkom/Siemens and Cell-Life.

References

[1] Commission of the European Communities. 2004. e-Health - making healthcare better for European citizens: An action plan for a European e-Health Area. Available at http://eur-lex.europa.eu/LexUriServ/site/en/com/2004/com2004_0356en01.pdf
[2] Accenture. 2006. Insights into the Introduction of Electronic Health Records in South Africa. Available at http://www.accenture.com/Countries/South_Africa/Research_and_Insights/EHRSouthAfrica.htm
[3] Jones, T. M. 2004. National Infrastructure for eHealth: Considerations for Decision Support. Studies in Health Technology and Informatics, Volume 100, 2004. E-Health - Current Situation and Examples of Implemented and Beneficial E-Health Applications. ISBN 978-1-58603-448-1
[4] The Department of Health and Human Services, USA. 2007. Summary of the NHIN Prototype Architecture Contracts: A Report for the Office of the National Coordinator for Health IT. Available at www.hhs.gov/healthit
[5] King, J. L. 1983. Centralized versus decentralized computing: organizational considerations and management options. *ACM Comput. Surv.* 15, 4 (Dec. 1983), 319-349. DOI= http://doi.acm.org/10.1145/289.290
[6] Ringholm. 2007. The Spine, an English national programme. http://www.ringholm.de/docs/00970_en.htm
[7] Puigjaner, R. 2003. Performance modelling of computer networks. In *Proceedings of the 2003 IFIP/ACM Latin America Conference on Towards A Latin American Agenda For Network Research* (La Paz, Bolivia, October 03 - 05, 2003). LANC '03. ACM, New York, NY, 106-123. DOI= http://doi.acm.org/10.1145/1035662.1035672
[8] OMNett++ simulation tool, http://www.omnetpp.org/
[9] A. Hanemann, A. Liakopoulos, M. Molina, M. Swany, *A Study On Network Performance Metrics and Their Composition*, Campus-Wide Information Systems (Emerald), Vol.23, No4, 2006.
[10] Bouzeghoub, M. 2004. A framework for analysis of data freshness. In *Proceedings of the 2004 international Workshop on information Quality in information Systems* (Paris, France). IQIS '04. ACM, New York, NY, 59-67. DOI= http://doi.acm.org/10.1145/1012453.1012464