December 2001

A New Method for Controlling Information Technology Hardware Complexity

S. Maj
Edith Cowan University

A. Boyanich
Edith Cowan University

Dasgf Veal
Edith Cowan University

A NEW METHOD FOR CONTROLLING INFORMATION TECHNOLOGY HARDWARE COMPLEXITY

P. Maj
Department of Computer Science
Edith Cowan University
p.maj@cowan.edu.au

D. Veal
Department of Computer Science
Edith Cowan University

A. Boyanich
Department of Computer Science
Edith Cowan University

Abstract

Knowledge of information technology hardware is an essential foundation for curriculum in Information Systems. However computer and network technology is not only a complex field but is also subject to many technical developments. These characteristics place considerable demands on teaching these subjects. However complexity may be controlled by the use of abstract models. This paper describes a new high-level abstract model called B-Nodes designed to control detail by means simple diagrams allowing top-down recursive decomposition. All work to date indicates that B-Nodes provide the basis of a new pedagogical framework that assists in the development of understanding, support different levels of technical detail and is valid for both current and future generations of digital technologies.

Introduction

The model curriculum and guidelines for undergraduate degree programs in IS (IS'97) clearly indicates that computer system hardware, networking and telecommunications are significant areas in IS curriculum. However computer and network technology is not only a complex field but is also subject to many technical developments. These characteristics place considerable demands on teaching these subjects. According to Davis, *'All aspects of the computing field have had rapid continuous change. As a result, university level Information Systems (IS) curricula need frequent updating to remain effective.'* (Davis et al, 1997). By example, PC design and manufacture has changed dramatically in the last decade. The trend towards Assembly Level Manufacturing (ALM), the use of Application Specific Integrated Circuits (ASIC's) and Surface Mount Technology (SMT) allow substantial manufacturing cost benefits and have resulted in a standard architecture with a modular construction. A state-wide survey in Western Australia found that computer technology is now managed as a modular system that demands skills other than those provided by traditional computer science curriculum (Maj et al, 1998). From this survey a set of guidelines were developed for the type of skills expected of computer science graduates entering the field of Information Technology (acquisition, deployment and management). Using the criteria developed a random selection of ten, successful, final year Edith Cowan University (ECU) computer science undergraduates were interviewed from a graduating population of approximately one hundred. The computer science degree at ECU is level one accredited, the highest, by the Australian Computer Society (ACS). The ACS curriculum is comparable to the ACM/IEEE 1991 Computing Curriculum. According to Maj none of the students interviewed had the skills expected by prospective employers. A parallel study of undergraduate and post-graduate Information Systems Management students at ECU gave comparable results. Thirteen students participated in this survey. Some of the postgraduate students interviewed already held managerial positions and were upgrading their qualifications. Others were engaged in employment and wished to move into IT management. A number of those interviewed either was, or had been, employed in the IT area. The initial ECU student questionnaire, first used in 1993, was also conducted in 1999 at two universities within the UK. Both universities have well established degree programs that are British Computer Society (BCS) accredited. The results obtained were directly comparable to the original survey (Maj et al, 2000). From these results it would appear that the current computer and network technology curriculum fails to meet the needs of potential employers. We therefore analysed this curriculum in detail.

Computer & Network Technology Curriculum

Professional bodies are responsible for curriculum content and accreditation. IS'97 provides the following course descriptions: *Information Technology Hardware and Software* (IS'97.4) and *Networks and Telecommunications* (IS'97.6). IS'97.4 recommends a range of hardware topics. Computing Curricula 1991 (CC'91) recommended nine subject areas that include *Architecture* as a knowledge unit. Architecture consists of a pre-requisite chain of topics that includes: digital logic, digital systems, machine level representation of data, assembly level machine organization, memory system organization, interfacing and communication and alternative architectures.. The 'Iron man' draft of Computing Curricula 2001 core body of knowledge also includes *Architecture and Organization* but in recognition of the importance of networking *Net-Centric Computing* is also now considered as a core body of knowledge. Both IS'97 and CC'91 are comparable in content with respect to their treatment of computer technology. Furthermore, both curricula recognise that computer technology can be described using a progressive range of models based on different levels of detail e.g. semiconductors, transistors, digital circuits. Such models are designed to progressively hide, and hence control detail, and yet provide sufficient information to be useful for communication, design and documentation. This is in keeping with the ACM/IEEE Computing Curricula 1991 in which abstraction is a recurring concept fundamental to computer science. By example, digital techniques and modelling provide an abstraction that is independent of the underlying details of semiconductor switching. Such digital circuits can be described without the complexity of their implementation in different switching technologies e.g. TTL, CMOS, BICMOS etc. Similarly details of semiconductor switching may be modelled using abstractions independent of the underlying details of quantum mechanics. Clements (Clements, 2000) makes the point that, '*… the generation of students studying electronics in the 1950's leaned about the behaviour of electrons in magnetic fields. The next generation studied transistor circuits, and the one after that studied integrated circuits. The traditional logic course changes rapidly*.' However, as discussed above, computer design and manufacture has changed rapidly in the last decade. Such developments have resulted in changes in the way in which the computer infrastructure is now perceived and managed, According to Clements, '*While the knowledge base academics must teach is continually expanding, only a fraction of that knowledge can be taught during a student's time at a university. Consequently, academics must continually examine and update the curriculum (computer architecture), raising the level of abstraction.*' We attempted therefore to develop, if possible, a new higher-level abstract model of computer technology. This model must represent a common conceptual framework held by students from different disciplines, and hence form the basis of a cognitive structure.

Modelling – Principles and Practices

Models are used as a means of communication and controlling detail. They should have the following characteristics: diagrammatic, self-documenting, easy to use, control detail and allow hierarchical top down decomposition. According to Cooling (1991), there are two main types of diagram: high level and low level. High-level diagrams are task oriented and show the overall system structure with its major sub-units. Such diagrams describe the overall function of the design and interactions between both the sub-systems and the environment. The main emphasis is 'what does the system do' and the resultant design is therefore task oriented. According to Cooling (1991), '*Good high-level diagrams are simple and clear, bringing out the essential major features of a system'*. By contrast, low-level diagrams are solution oriented and must be able to handle considerable detail. The main emphasis is 'how does the system work'.

Computer technology can be modelled using symbolic Boolean algebra (NOR, NAND gates). These gates may be implemented using solid-state electronic switches or even gas state electronics i.e. thermionic valves. At this lower level, the basic implementations of solid state switching may be described with models directly relevant to engineers at this level of operation. Logic gates may be connected to create combinatorial and sequential circuits and hence functional units such as Read Only Memory (ROM) etc. Such functional units can also be modelled but using high level diagrams. The underlying switching technology is not relevant at this higher level of abstraction. At an even higher level of abstraction computer technology can be modelled as a collection of programmable registers. However, whilst useful, none of these models are directly relevant to technical details such as hard disc speeds, local bus performance etc. As suggested above, the PC is now a low cost consumer item with a standard, heterogeneous architecture and modular construction. A higher level of model is therefore needed that is directly relevant to this current technology.

B-Nodes

It is a common experience to perceive the PC as a modular device (CDROM, Zip Drive, Modem etc) used to store, view and process either local or networked data. The traditional bottom up method of teaching computer technology is not a good constructivist approach. Accordingly Scragg recommends a top down approach starting with material already familiar to students

and then working towards less familiar models (Scragg, 1991). Individual modules in a PC (microprocessor, hard disc drive etc), and the PC itself, may be modelled using B-Nodes (Maj et al, 2000a). Each B-Node can be treated as a data source/sink capable of, to various degrees, data storage, processing and transmission. The performance of each B-Node may be calculated, to a first approximation, by Bandwidth = Clock Speed x Data Path Width (B = C x D) with units in either MBytes/s or Frames/s. A frame is defined as 1024x1024 pixels with a colour depth of 3 bytes per pixel i.e. 3MBytes. This simple, high-level, task oriented model may provide a suitable conceptual map and hence the framework for an introduction to computer technology. Even though technical detail is lost, this model is conceptually simple, controls detail by abstraction and may allow students to easily make viable constructs of knowledge based on their own experience. The units Frames/s may be more meaningful to a typical user because it relates directly to their perception of performance. To a first approximation, smooth animation requires approximately 30 Frames/s (90MBytes/s). According to Barney, '*A measurement is the process of empirical objective assignment of numbers to properties of objects or events in the real world in a way such as to describe them'* (Barney, 1985). History has many examples of measures in the search for useful standards. Early Egyptians defined one finger-width as a zebo and established an associated simple, reproducible and denary scale of standard measurements. It is significant that human dimensions were used as the basis of one of the first standards. If B-Nodes are not used hardware selection is based on a wide range of units (Microprocessor -MHz, Electronic memory - nanosecond, Hard Disc Drive - rpm, CDROM – speed etc). Evaluation of these heterogeneous modules, each using different units of measurement, is therefore difficult. B-Nodes can be used to model heterogenous sub-modules within a PC (microprocessor, hard disc drive, bus structures, network etc) using simple, meaningful, derived units with a denary scale. We have therefore a common unit of measurement, relevant to common human perception, with decimal based units, that can be applied to different nodes and identify performance bottlenecks. The use of simple, fundamental units allows other units such as frame transfer time to be easily calculated. This allows the performance of heterogeneous units to be directly compared (Table 1) using the same units.

Table 1. Bandwidth

Device	Clock Speed (MHz)	Data Width (Bytes)	Bandwidth (MBytes/s) B = C x D	Bandwidth (Frames)	Frame transfer time (1/Frames)
Processor	400	8	3200	1066	0.9ms
DRAM	16 (60ns)	8	128	42	23ms
Hard Disc	60rps	90Kb	5.4	1.8	0.6s
CROM	(30 speed)	(150Kbytes/s)	4.6	1.5	0.6s
ISA Bus	8	2	16	5.3	0.18s
Ethernet	100	1/8	12.5	4.1	0.24s

The characteristics of the Frame may be changed to more directly suit different applications. By example medical images are often stored digitally. A single ultrasound image represents approximately 0.26MBytes of data (Dwyer, 1992). The performance of each B-Node may be calculated using this metric. The use of B-Nodes has been confirmed experimentally (Maj and Veal, 2000b). The B-Node model has been successfully applied to a wide range of PC architectures allowing a direct comparison not only between different B-Nodes within a given PC but also comparisons between different PC's. Using B-Nodes it was possible to analyse PC's with different Intel microprocessors (8088/6, 286, 386, 486 etc.) and various associated bus structures (Micro Channel Architecture, Extended Industry Standard Architecture, Video Electronic Standards (VESA) Local Bus).

Sub-optimal Operation

B-Nodes typically operate sub-optimally due to their operational limitations and also the interaction between other slower nodes. For example, a microprocessor may need two or more clock cycles to execute an instruction. Similarly a data bus may need multiple clock cycles to transfer a single data word. The simple bandwidth equation can be modified to take this into account i.e. Bandwidth = Clock x Data Path Width x Efficiency (B = C x D x E). The early Intel 8088/86 required a memory cycle time of 4 clocks cycles (Efficiency = ¼) however, for the Intel 80x86 series, including the Pentium, the memory cycle time consists of only 2 clocks (Efficiency = ½) for external DRAM. Efficiencies can be calculated for each device and the performance calculated accordingly (Table 2). However, other factors not considered include the effects of compression, operating system overheads etc. The effect of these is currently being examined.

Table 2. Bandwidth with Efficiency

Device	Clock Speed (MHz)	Data Width (Bytes)	Efficiency	Bandwidth (MBytes/s) B = C x D x E	Bandwidth (Frames/s)
Processor	400	8	0.5	1600	533
DRAM	16 (60ns)	8	0.5	64	21
Hard Disc	60rps	90Kb	0.5	2.7	0.9
CROM	(30 speed)	(150kBytes/s)	0.5	2.3	0.8
ISA Bus	8	2	0.25	4	1.3
Ethernet	100	1/8	0.9	11.25	3.8

B-Nodes and E-Business Architecture

Service Level Agreements in conjunction with cost constraints and the technologies used are the primary determinants in performance and capacity planning. In this context a variety of models are used. The Business model defines the purpose of the organization; the Functional Model defines the navigational structures and the Customer Model is used to describe the user behaviour patterns. Using the customer model, the number of clients, type of resources requested, pattern of usage etc are all used to determine the workload characteristics. Workload characteristics, in conjunction with the resource infrastructure model will determine site performance and whether or not the Service Level Agreements can be met. Customer Behaviour Modelling methods have been successfully used to determine aggregate metrics for E-Commerce web sites (Menasce et al, 1999). Using these various models it is possible to obtain a wide variety of different performance metrics that include: Hits/s, Page Views/Day, Unique Visitors etc. However there are problems with using these metrics to define the characteristics of the required infrastructure. There appears to be no simple method to convert these various metrics to units that can be directly used to evaluate hardware performance. If a web server is modelled as a B-Node then the performance metric is bandwidth with units of Mbytes/s. The sub-modules of a server (microprocessor, hard disc, electronic memory etc) and also be modelled as B-Nodes, again using the same performance metric. The use of fundamental units (Mbytes/s) allow other units to be derived and used e.g. transactions per second (tps). Assuming the messages in a client/server interaction are 10kbytes each, the performance of each B-Node can be evaluated using the units of transactions/s (Table 3)

Table 3. Bandwidth (Transactions)

Device	Clock Speed (MHz)	Data Width (Bytes)	Efficiency	Bandwidth (MBytes/s) B = C x D x E	Bandwidth (Transactions)	Load (Transactions)	Utilization
Processor	400	8	0.5	1600	160k	250	<1%
DRAM	16 (60ns)	8	0.5	64	6.4k	250	4%
Hard Disc	60rps	90Kb	0.5	2.7	270	250	93%
CROM	(30 speed)	(150kBytes/s)	0.5	2.3	230	250	>100%
ISA Bus	8	2	0.25	4	400	250	63%
Ethernet	100	1/8	0.9	11.25	1.1k	250	23%

If the demand on this server is 250 Transactions/s it is a simple matter to determine both performance bottlenecks and also the expected performance of the equipment upgrades. From table 3 it is possible to determine that for this web server, the hard disc drive, CDROM and ISA bus are inadequate. The metric of transactions/s can easily be converted to the fundamental unit of Mbytes/s, which can then be used to determine the required performance specification of alternative bus structures, CDROM devices and hard discs. A PCI (32 bit) bus structure is capable of 44Mbytes/s. A 40-speed CDROM device has a bandwidth of approximately 6Mbytes/s. Similarly replacing the single hard disc drive by one with a higher performance specification (rpm and higher track capacity) results is a new server capable of meeting the required workload (Table 4).

Table 4. Upgraded Server

Device	Clock Speed (MHz)	Data Width (Bytes)	Efficiency	Bandwidth (MBytes) B = C x D x E	Bandwidth (Transactions)	Load (Transactions)	Utilization
Processor	400	8	0.5	1600	160k	250	<1%
DRAM	16 (60ns)	8	0.5	64	6.4k	250	4%
Hard Disc	100rps	250K	0.5	12.5	1.25k	250	20%
CROM	(40 speed)	(150kBytes/s)	0.5	6	0.6k	250	42%
PCI Bus	33	4	0.5	66	6.6k	250	4%
Ethernet	100	1/8	0.9	11.25	1.1k	250	23%

Capacity planning is the process of predicting future workloads and determining the most cost-effective way of postponing system overload and saturation. Assuming that the web traffic is anticipated to rise to 550 transactions/s – the current single server solution will be inadequate. To accommodate much higher web traffic a typical e-business configuration may consist of a front-end Web server, a Secure Web server, a Payments server, an Application server and a Database server. Assuming each server is a separate device connected by a 100Mbps Ethernet link it is possible to model this configuration using B-Nodes. Each server represents a B-Node. The communication link may be represented as a directed arc (arrow) annotated by its bandwidth performance (units MBytes/s or Transactions/s). Using Customer Behaviour Model Graphs (CBMG's) it is possible to evaluate the relative frequency that each dedicated server is used. Assuming probability based on relative frequency our performance equation is now Bandwidth = Clock Speed x Data Path Width x Efficiency x Frequency (B = C x D x E x F). For each server the results can be tabulated (table 5). The load data obtained from table 5 can then be used to evaluate the performance of the individual components in each server. In the case of the Web server the actual load is 5.625Mbyte/s (0.55kTransactions/s) from which the utilization of each module can be evaluated (table 6).

Table 5. E-Commerce Servers

Device	Network Bandwidth (Mbytes/s)	Network Bandwidth (Transactions/s)	Relative traffic frequency (f)	Actual server load (Mbytes/s)	Actual server load (Transactions/s)
Web server	11.25	1.1k	0.5	5.625	0.55k
Secure server	11.25	1.1k	0.05	0.5625	0.055k
Payment server	11.25	1.1k	0.05	0.5625	0.055k
Database server	11.25	1.1k	0.2	2.25	0.22k
Application server	11.25	1.1k	0.1	1.125	0.11k

The CMBG's clearly indicate that the majority of the traffic is to the Web server. Assuming that it is necessary to plan for an expected load of 1,000 transactions/s. From table 6 it is evident that the web server would not perform satisfactorily. Traffic characterization may be used. Assuming that such an analysis indicates that 60% of the traffic is for static JPEG images and 40% for dynamic HTML pages. A possible solution may be to use a caching proxy to serve the static web pages.

Table 6. Evaluation of E-Commerce Web Server

Web Server	Clock Speed (MHz)	Data Width (Bytes)	Efficiency	Bandwidth (Mbytes) B = C x D x E	Bandwidth (Transactions)	Load (Transactions)	Utilization
Processor	400	8	0.5	1600	160k	550	<1%
DRAM	16	8	0.5	64	6.4k	550	9%
Hard Disc	100rps	250K	0.5	12.5	1.25k	550	44%
CROM	(40 speed)	(150kBytes/s)	0.5	6	0.6k	550	92%
PCI Bus	33	4	0.5	66	6.6k	550	8%
Ethernet	100	1/8	0.9	11.25	1.1k	550	50%

Assuming that the traffic analysis further finds that this brings down the average message size to the server from 10Kbytes to 5kBytes – the web server can easily be modelled using B-Nodes accordingly. For the web server 40% of 1,000 transactions/s is 400 transactions/s, furthermore each transaction is on average 5kBytes. This results in a different, and much lower, utilization for the Web server (Table 7).

Table 7. Evaluation of E-Commerce Web Server

Web Server	Clock Speed (MHz)	Data Width (Bytes)	Efficiency	Bandwidth (Mbytes) B = C x D x E	Bandwidth (Transactions)	Load (Transactions)	Utilization
Processor	400	8	0.5	1600	320k	400	<1%
DRAM	16	8	0.5	64	12.8k	400	3%
Hard Disc	100rps	250K	0.5	12.5	2.5k	400	16%
CROM	(40 speed)	(150kBytes/s)	0.5	6	1.2k	400	33%
PCI Bus	33	4	0.5	66	13.2k	400	3%
Ethernet	100	1/8	0.9	11.25	2.2k	400	18%

B-Nodes as a Pedagogical Framework

As a result of the initial investigations at ECU a new curriculum based on B-Nodes was designed, implemented and fully evaluated at ECU. Unlike the standard computer technology curriculum students are not taught digital techniques, assembly language programming etc. This curriculum has always been oversubscribed, has a very low student attrition rate, and attracts students from other Faculties in ECU and students from other universities in the state. Using the standard compulsory ECU course evaluation questionnaire the unit was highly rated by students. Furthermore, a more detailed study was conducted to investigate student experience of the B-Node concept. From an enrolment of eighty students, forty were randomly selected and given questionnaires. Thirty-six students thought the B-Node concept should be taught. Thirty-five students thought that this concept helped them understand computer technology. Thirty-five students thought that using a common unit (Frames/s) helped in evaluating PC devices.

Conclusions

This paper proposes B-Nodes, whose performance is rated by bandwidth (MBytes/s), as a method for modelling computer and network systems. B-Nodes represent a new, higher level of abstraction that allows technical detail to be controlled using top, down recursive decomposition. Given this model provides abstraction it is therefore independent of architectural detail and can therefore accommodate rapid changes in technology. It is valid for all generations of digital PC technology to date and may therefore continue to be useful for some years to come. The use of fundamental units allows other, more useful, units to be derived (Frames/s, Transactions/s). The use of fundamental units (Mbytes/s) allows the performance of heterogeneous devices to be directly compared and evaluated. Furthermore, other more meaningful units (e.g. frames/s, transactions/s etc) may be derived. Derived units may be used to evaluate the performance of an e-commerce web site for capacity planning. Significantly as these

units are derived it is a simple task to convert these units into fundamental units that can be directly used for the evaluation and selection of hardware. Using B-Nodes it is possible to model not only the modules within a PC but also network and E-Business client/server architecture. Work to date strongly indicates that B-Nodes can be used as the pedagogical basis of curriculum in computer and network technology.

References

Barney, G. C.. Intelligent Instrumentation - Microprocessor Applications in Measurement and Control. Exeter: Prentice/Hall International (UK), 1985.

Clements, A. "Computer Architecture Education", *IEEE Micro*, Vol. May-June, 2000, pp. 10-22.

Cooling, J. E. Software Design for Real-Time Systems. Padstow, Cornwall: Chapman and Hall, 1991.

Davis, G. B., Gorgone, J. T., Couger, J. D., Feinstein, D. L., & Longenecker, H. E. IS'97. Model Curriculum and Guidelines for Undergraduate Degree Programs in Information Systems: Association of Information Technology Professionals, 1997.

Dwyer, S. Teleradiology Using Switched Dialup Networks. IEEE Journal on Selected Areas in Communication Sept., 1997.

Maj, S. P., Robbins, G., Shaw, D., & Duley, K. W. "Computer and Network Installation, Maintenance and Management - A Proposed New Curriculum for Undergraduates and Postgraduates". *The Australian Computer Journal*, 30(3), 1998, pp. 111-119.

Maj, S. P. & Veal, D. "Computer Technology Curriculum - A New Paradigm for a New Century". *Journal of Research and Practice in Information Technology*, 32 August/September, pp. 200-214,2000a.

Maj, S. P. & Veal, D. "Architectural Abstraction as an aid to Computer Technology Education". Paper presented at the American Society for Engineering Education (ASEE), St Louis, Missouri, 2000b.

Maj, S. P., Veal, D., & Charlesworth, P. "Is Computer Technology Taught Upside Down?", Paper presented at the 5th Annual SIGCSE/SIGCUE Conference on Innovation and Technology in Computer Science Education, Helsinki, Finland, 2000.

Menasce, D. A., Almeida, V. A. F., Fonseca, R. C., & Mendes, M. A. "A Methodology for Workload Characterization for E-Commerce Servers". Paper presented at the 1999 ACM Conference in Electronic Commerce, Denver, CO, 1999.

Scragg, G. W. "Most computer organisation courses are built upside down". *ACM SIGCSE Bulletin*, 23(1), 1991. pp. 341-346.