Volume 10 Issue 2 November 1995

ICL

Ingenuity

THE TECHNICAL JOURNAL

Ingenuity™

ICL's Technical Journal

All correspondence and papers to be considered for publication should be addressed to the new Editor.

The views expressed in the papers are those of the authors and do not necessarily represent ICL policy.

Published twice a year by Multimedia Publishing Unit, ICL, Bracknell.

1996 subscription rates (including postage & packing):

	UK and Europe	Rest of World
Annual subscription	£72	$144
Single issues	£43	$86

ISSN 1354-9952

Ingenuity

Volume 10 Issue 2

Contents

Other services

Front cover: Richard Branson and Per Lindstrand with their engineering and logistics team in Japan prior to the flight of the Virgin Trans Pacific Project, January 1991.

Editorial

From the beginning of 1996, the editor of the Journal will be V.A.J. Maller, ICL Professor of Computer Systems at Loughborough University. He will also be known to many in ICL as a member of Gordon Scarrott's research team; he has had much research experience outside ICL since.

A retiring editor may be allowed a certain licence when contributing his last editorial which I mean to exploit by looking both backwards and forwards. Reminiscences should be brief and lead to some conclusion as to possible future trends. I can recall the twenties, the era of crystal radio sets, plug in coils and bright emitter valves. At school in the thirties I built many radios of increasing complexity, the last an eight valve, four wave-band superhet with a push pull triode output stage, gramophone pick up and some pretension to high fidelity. At that time computers had been heard of by very few people (Zuse, who built an electromechanical computer in Germany, was an exception).

During the war I worked for the government on RAF radar systems, chiefly for IFF (a radar identification system), and had experience of test flights, writing specifications and supervising contractors. I entered the computer industry in 1949 before it really existed, to work for J. Lyons and Company thanks to having done a PhD in ultrasonics and (I suspect) a good reference from Professor Maurice Wilkes. The design, construction and operation of the LEO series by a small team followed and I survived two mergers to end up in ICL from which I retired in 1984. I was extremely fortunate to have had a richly varied series of jobs in both design and management applied to computer hardware architecture, in IT standards through ECMA and finally in marketing.

Throughout computer history there have been problems in relating human, mechanical and various electronic time scales. Direct keying in of input data was many orders of magnitude too slow to satisfy the computer's appetite for data to process. Human reactions to reports of difficulties with data or processing were and still are very slow by comparison with computer behaviour. Input and output devices themselves were slow and could keep processing waiting. So could backing stores of every kind. Even direct access to main store was still slow compared with processing leading to many ingenious forms of buffering and caching.

These problems are still with us today and the compromises adopted to resolve them seem unlikely to be stable. The ways we have found to overcome them all involved numbers of people or a multiplicity of devices being set to work in parallel with and/or ahead of the moment data or programs were required. With batch processing for example one can normally see far enough ahead to get a succession of data elements ready to be processed. With transaction processing, however, prediction of what data or enquiries are to come next can be only statistical. Problems of splitting up complex programs into pieces that allow separate teams of programmers to be deployed in parallel, seem hard even to state. Perhaps the human mind is not well adapted to visualising and then programming multiple operations in parallel.

But the balance of advantage is constantly shifting as differing techniques of processing, storage and communication advance out of step with one another. Today there is debate over whether everyone needs a complete PC with its elaborate operating system, private discs and optical storage to do many of the things it is being used for now. The speed of optical fibre communication makes it conceivable to satisfy hundreds or even thousands of users from a single server. This could enormously reduce the labour of installing and constantly updating application and system software, a task otherwise faced by every user individually. Clearly not all classes of users or of uses can be satisfied in such a way: the issue is whether there is a potentially profitable market for this form of working.

It seems certain increased speed of communication will lead to it becoming far cheaper, and usefully quicker, to send data in bulk, though not necessarily so much cheaper to send small packets at random instants in time. Forecasting the overall effect of technological advances on exactly how people and organisations will come to use IT becomes ever more problematic: current uses may remain but new ones will certainly be added, especially where the man or woman in the street can originate a task directly. Less prior knowledge and less skill will be demanded of the ultimate end user. Kiosks such as are featured in this issue are an excellent example of this trend.

Few can claim to have foreseen the invention of the transistor more than 40 years ago or, 30 years ago, the potential of the levels of circuit integration now possible on one chip. The full impact of digital representation of data, sound, images (and of almost every other form of communicable intelligence) remains unimagined. Perhaps the only safe conclusion from looking both backwards and forwards is that one is sure to be surprised by events.

In conclusion, I wish the new editor all possible success in building on the traditions so firmly established by my predecessor Dr Jack Howlett, the founder of the ICL Technical Journal. In this Vic Maller is sure of the enthusiastic support of the Editorial Board. May all readers commend the Journal to their friends and colleagues both inside and outside ICL.

The Architecture of the ICL GOLDRUSH MegaSERVER[1]

Paul Watson[2] and George Catlow

ICL, High Performance Technology,
West Gorton, Manchester M12 5DR, UK

Abstract

This paper discusses the requirements which are to be met by a parallel computer system if it is to satisfy the requirements of commercial database processing, and describes how one such system – the ICL GOLDRUSH MegaSERVER – has been designed to meet these requirements.

GOLDRUSH is a distributed store parallel processor consisting of up to 64 Elements, each of which can co-operate in database processing, exploiting both the parallelism found within complex queries (intra-query parallelism) and that found between queries in On-Line Transaction Processing workloads (inter-query parallelism). The paper discusses the requirements of business critical database applications including high availability, integrity and manageability. It then details the architecture of GOLDRUSH in order to show how a commercially available system has been designed to meet these requirements; these include resilience to failure of hardware components such as disks and processors, and the provision of system management applications which allow the parallel machine to be managed as a single system.

1. Introduction

Conventional mainframe computers are being pushed to, and beyond, their performance limits by the needs of today's commercial computing workloads.

---

[1] This paper is a version of that read at the British National conference on Databases in July 1995, published in the conference proceedings: Lecture Notes in Computer Science 940, Advances in Databases, ed. C. Goble and J. Keane, Springer-Verlag, 1995, ISBN 3-540-60100-7.

[2] Current Address: Department of Computing Sciences, University of Newcastle, Newcastle-upon-Tyne, NE1 7RU, UK.

This, combined with the cost of these large systems, is forcing both customers and manufacturers to look at more radical architectural alternatives. Parallel computers have been used successfully for many years to overcome these same problems for scientific computing [Alamasi & Gotlieb, 1989], but it is only now that parallel machines are starting to be produced for the commercial market by a range of mainstream computer manufacturers. Potentially, this market could become far larger than the scientific market: 80% of the world's large computer systems (those over $1M) are commercial rather than scientific [International Data Corporation, 1992].

As these new commercial parallel systems become available, the unprecedented levels of performance they offer will create new opportunities for businesses which utilise them. Already companies are re-engineering their business processes to fully exploit the new technology. One example of this is a financial organisation with a large number of customers, each of whom has a number of different types of account. Currently, information on each account is held in a different database, on a different computer. This makes it impossible for the organisation to collate information about all the accounts held by a given customer. They wish to do this so as to form an overall picture of the financial state of their customers, allowing them to target their services and marketing more effectively. The solution is to move to a parallel system: the higher levels of performance offered will allow them to unify all their information into a single database, on a single machine. The database can be structured to allow information to be accessed by both account number and customer.

The GOLDRUSH system evolved from earlier work carried out in collaborative projects with other companies and Universities. The two most important of these projects were: the Alvey Flagship project [Watson, et al, 1988] which produced a distributed store parallel system running both Declarative Languages and Databases; and the ESPRIT EDS project [Watson & Townsend, 1991] which developed a parallel system for Parallel Databases, Language Translation and Declarative Languages.

The rest of this paper focuses on the GOLDRUSH parallel system, and is structured as follows. The requirements of commercial parallel database systems are discussed in section 2. Section 3 then describes the overall GOLDRUSH architecture, including the hardware. Section 4 details the platform software and 5 describes how databases are supported by it. Sections 6 and 7 give details of how two of the key requirements for commercial parallel computers – high availability and manageability – are met. Finally, section 8 contains some of the conclusions we have drawn from our work.

2. Requirements

The challenge for the designers of parallel machines is to give customers the key attributes they have traditionally found in conventional mainframes, but with greater power and lower cost-performance. Customers will not entrust their business-critical applications and data to a machine which does not offer high availability and manageability. The methods which have been adopted to meet these requirements in GOLDRUSH are described in sections 6 and 7.

Another attraction of parallel machines is their scalability; customers can grow their systems by adding units of processing power and IO capability as required. It is important that the design supports this by ensuring that there are no system bottle-necks which reduce scalability.

The need to run commercial relational database systems (RDBMS) places two major requirements on the platform software of the machine. Firstly, there is a need for a globally shared filestore which all processors in the system can access so that each can be running transactions against the same database tables. Secondly, a Global Lock Manager is required to preserve the integrity of shared database tables when multiple processors are performing transactions simultaneously. The design of these two components is described in section 4.

It is interesting to note that as yet there is no requirement for distributed shared store: current commercial parallel database systems attain parallelism by running a separate server on each processor. They communicate only through the global lock manager, and the globally shared filestore, both of which can be efficiently implemented by distributed, message passing processes.

The implementation of the lock manager and filestore in GOLDRUSH is described in section 4. The mapping of the database servers onto the system is described in more detail in section 5.

3. Overall Architecture and Hardware

The GOLDRUSH MegaSERVER is a Database Server, designed to run commercial database back-ends. It stores the database, and services SQL queries sent by external clients such as PCs, workstations and mainframes. An example application architecture is shown in figure 1.

The diagram shows options for Management Information Systems (MIS) generating complex queries, On Line Transaction Processing (OLTP) and Batch Clients. The use of Transaction Processing (TP) protocols in the application architecture is recommended for large systems: clients connect over a local or wide area network to an Application server, which generates the SQL. This improves performance as TP sends less data over a wide area network than does SQL; it also offers better security and control of application code. Note that various existing systems can be connected to GOLDRUSH. The applications can remain on these systems, but the database migrates to GOLDRUSH.

The internal architecture of a GOLDRUSH system is shown in figure 2. It consists of a set of Processing Elements (PEs), Communications Elements (CEs) and Management Elements (MEs) connected together by a high performance network (DeltaNet).

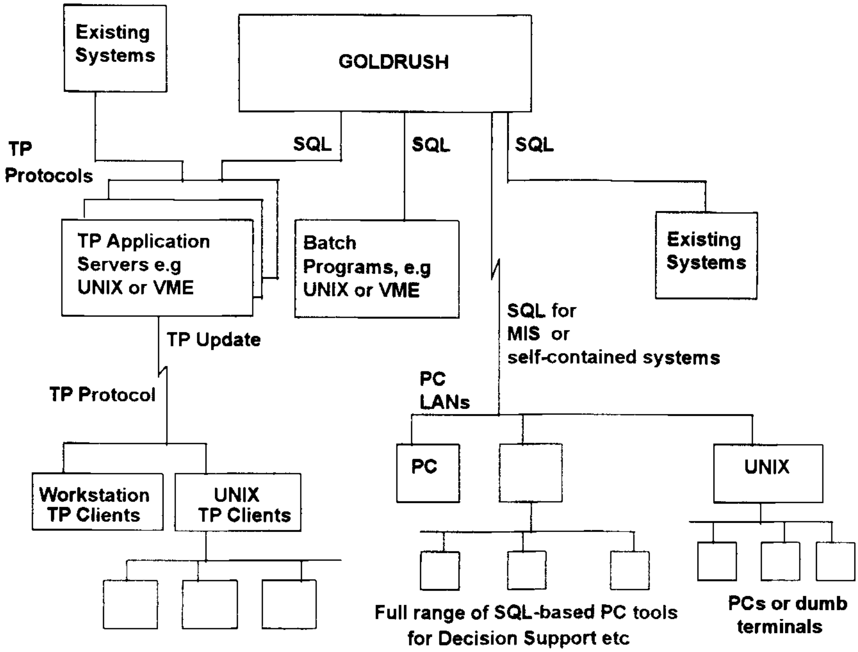


Figure 1 Example GOLDRUSH Application Architecture

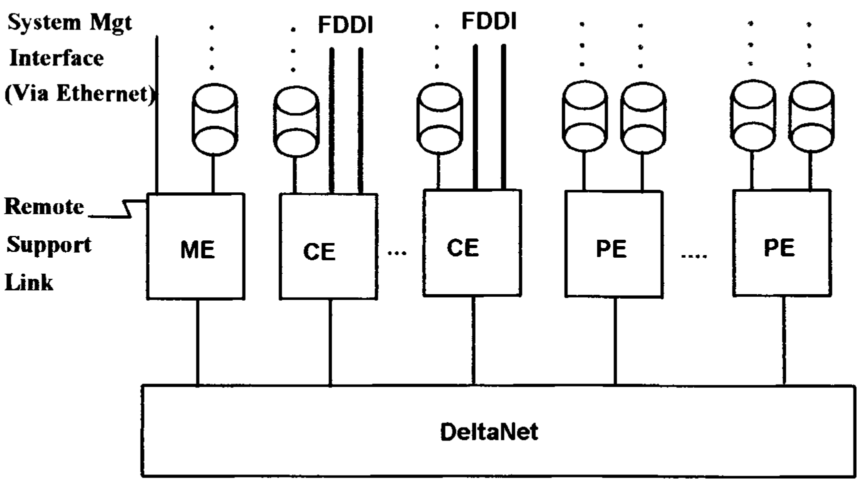


Figure 2 The Internal GOLDRUSH Architecture

The PE is the most common Element and is designed to run a Database Back-end. It consists of two SPARC RISC microprocessors, one of which runs the database server while the other is dedicated to delivering high performance message passing over the DeltaNet. A very large amount of RAM store is provided in the PE to enable large database caches to be configured: this is very important for achieving high performance database processing. Each PE also has two SCSI-2 (each both wide and fast) connections, allowing up to 30 disks to be connected.

The Communications Element is identical to the Processing Element except that one of the SCSIs is replaced by two FDDI couplers for Client connection. FDDI was chosen because of its high performance (100Mb/s), and the availability of bridges allowing connection to all other common types of LAN and WAN, such as Ethernet. Multiple CEs can be configured in a system for both performance and resilience reasons (see section 6).

The Management Element is a conventional mid-range UNIX processor (currently an ICL DRS6000) which runs the Management Software (which is described in section 7). It also contains a "Teleservice" modem connection allowing: problem reports to be sent to a service desk; remote problem diagnosis from the service centre; and, software problem fixes to be sent from the service centre to GOLDRUSH.

The DeltaNet is a high performance Delta Network built out of 8x8 router chips. 128-Byte Messages are sent through the DeltaNet between Elements over full duplex links delivering up to 25MBytes per second each way per Element.

For high performance archiving, tape libraries can be attached to the Elements via their SCSI connectors. Multiple Element connections provide high performance archiving by allowing multiple data streams to be archived simultaneously to multiple tape drives.

GOLDRUSH systems can contain up to 64 Elements. Each PE and CE has two 90MHz HyperSPARC processors, with 256MBytes of RAM. PEs can each connect to up to 50GBytes of Disk Store. Because of the reliance on industry standards, commodity components, this specification will be continuously upgraded as new versions of the components become available. These upgrades will include faster processors and larger capacity disks. The first release offers the Ingress, Oracle, Adabas and Informix databases.

4. Platform Software

Figure 3 shows the GOLDRUSH platform software architecture. Each Processing and Communications Element runs a Chorus micro-kernel based SVR4 UNIX Operating System. This has been enhanced to support the requirements of parallel database processing by the addition of the following sub-systems:

4.1 Internal Communication Services

As described in the previous section, the DeltaNet hardware offers fast communication between Elements. In a parallel machine such as GOLDRUSH which relies entirely on message passing for inter-PE communications it is very important that low latency, high throughput message passing is available to software. While conventional transport protocols such as TCP/IP and ISO transport are available for inter-Element communications, they add an unacceptably large overhead on top of the basic hardware performance of the DeltaNet. In order to overcome this, a very lightweight communications protocol has been designed. It is made available through UNIX interfaces so that application level software can exploit it; for example database servers can use it for communicating fragments of queries when exploiting intra-query parallelism. Internal kernel interfaces are also provided, and these are used by a number of the platform sub-systems, for example for remote filestore access and distributed deadlock detection. The bulk of the lightweight comms stack runs on the SPARC processor dedicated to driving the DeltaNet (see section 3), so minimising the performance impacts on the other SPARC processor which runs the database server.

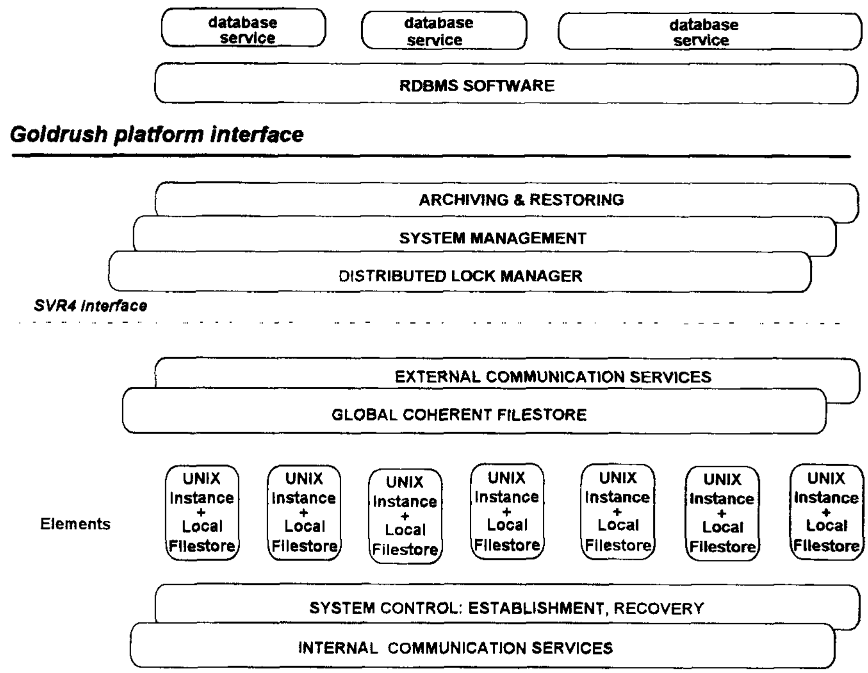


Figure 3 GOLDRUSH Platform Software Architecture

4.2 System Control

As was described above, each Element runs a separate instance of UNIX SVR4. It is a requirement that it should not be necessary to manage each UNIX instance separately. Therefore, it is the job of the system control layer to support this: it is responsible for establishing and shutting down the Elements in a co-ordinated fashion. The standard Network Information System (NIS) is used to provide a global method of configuring all the Elements, including users and communications connections. If Disks or Elements fail then this layer ensures that the rest of the system can continue and that applications remain running. This is discussed in detail in section 6.

4.3 Global Coherent Filestore

As described earlier, each Processing Element (PE) runs a separate database server. However, they are all able to perform transactions on the same database simultaneously. This requires that each PE can access all the database tables which are held on disk (details of the mapping of tables onto disks is given in section 5). As was described in section 3, each Processing Element contains two SCSIs for disk connection. Therefore, physically each disk is attached to one Element only. Some of the space on the disks is dedicated to local filestore which is only required to be accessible by that Element: for example the local UNIX kernel. However, the rest of the disk space is used to hold database data (tables, logs etc.) which must be globally accessible by other Elements if parallel database processing is to be supported. Providing this connectivity is the role of the Global Coherent Filestore (GCF). Each Element creates its own portion of this Global Filestore on its own local disks, and then cross-mounts the global filestore from all other Elements, so making it accessible to the database server running on it. In this, the GCF is very similar to the standard UNIX Network File System (NFS); however there are two important differences. Firstly, the GCF has been engineered to provide very high performance remote filestore access. To achieve this it is implemented in the kernel, and exploits the lightweight communications described above. Secondly, unlike NFS, the filestore is completely coherent. This ensures that if one Element updates part of the database, the updated value is seen by all other Elements which subsequently access that part of the database. This is achieved by implementing only server-side filesystem caching (NFS also has client-side caching). The lack of client-side caching does not reduce performance as the database servers themselves manage their own client-side cache in main store on each Element (which is why the large main stores are required).

For efficient database support, both Asynchronous and Raw IO are supported by the GCF, along with conventional UNIX file access. For high availability, the Veritas VxVM Volume Manager is used to allow all data to be mirrored, while for fast recovery after Element failure the Veritas VxFS Filesystem is used. Their use is described in more detail in section 6.

4.4 Global Communication Services

External Communication between GOLDRUSH and external Clients generating SQL is via the FDDI couplers on the CEs. Both TCP/IP and ISO transport protocols are supported.

It was an important criterion for manageability that GOLDRUSH appeared as a single system to clients, rather than as a set of individual Elements. The Global Communications Services layer provides this in the following way. For ISO communications the GOLDRUSH machine has a single externally visible address independent of the number of CEs and PEs. For TCP/IP communications each FDDI coupler has a separate address, independent of the number of PEs. PEs wishing to set up communications with clients "listen" for connections to the service they are offering, and when an external client attempts to connect to the service, the CE which receives the request routes it to a PE offering the service. If multiple PEs are offering the service then the CE will load-balance the connection requests across those PEs. Once a connection between an external client and a PE is established, the CE relays the data packets from the FDDI coupler to the PE (and vice versa) using the lightweight communications protocol for efficiency. The Global Communications Services layer can maintain connections in the presence of FDDI coupler, CE or (in some cases) external LAN failure. This is described in section 6.

4.5 Distributed Lock Manager

On GOLDRUSH, database management servers run in parallel on many PEs. It is the job of the Distributed Lock Manager to ensure that they run transactions consistently, avoiding concurrent, incompatible updates to the databases. On a uniprocessor, this job is done by the database server's own lock manager which receives requests to take and drop locks and forces transactions to wait until they have the correct lock before data can be accessed. The Distributed Lock Manager (DLM) must act as a Global Lock Manager for all the database servers running against the same database. It is distributed across all Elements, with an instance running on each Element, so that the processing and communications load is shared, and no one Element becomes a bottle-neck. Each lock is managed by one instance of the DLM and all requests for it are sent to that instance by the PE generating the request. The DLM also provides Global Deadlock detection using a novel algorithm to minimise inter-PE communication [Hilditch & Thomson, 1989]. For efficiency, all communication with the DLM is by the lightweight communications protocol. The DLM is also resilient to the failure of any process or Element (see section 6).

4.6 System Management

Section 7 describes this layer.

4.7 Archiving & Restoring

One of the key attributes found in mainframes but generally not available in UNIX-based solutions is fast archive and restore. This is important as the time taken by archiving may reduce the availability of the system for database processing. Also the time taken to restore from archive will limit the recovery time after a major failure. The latter can be a particular problem as a major failure, unlike archiving, cannot be planned in advance.

The large disk capacity of systems such as GOLDRUSH requires very high throughput rates. Therefore for very high performance, this layer supports the integral, SCSI connected archiving described in section 3, ensuring that data is archived as fast as the available tape drives can accept it. For lower performance archiving, which may for example be acceptable for small databases, the layer will support archiving to external archive servers connected via the FDDI connections.

5. Databases

The initial release of GOLDRUSH supports the Oracle, Ingress, Adabas and Informix databases. There are two basic mappings onto the GOLDRUSH architecture. These are denoted as the Shared Access and Distributed Access mappings and are described in this section.

The Shared Access mapping is shown in figure 4.

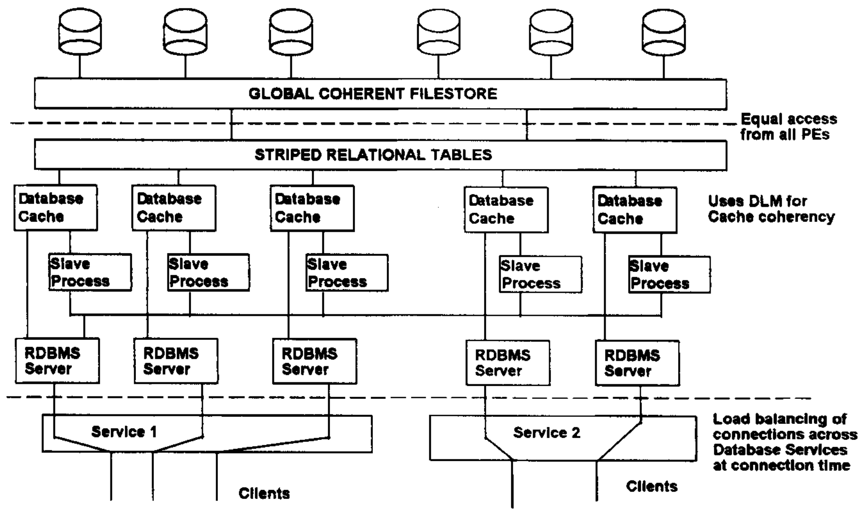


Figure 4 Shared Access Mapping of Databases onto GOLDRUSH

In this mapping, each PE runs its own Database Server. The database tables are striped across disks attached to a number of processing Elements; the Global Coherent Filestore allows all Elements to access all fragments of the tables. Each PE has a local database cache in main store, and the Distributed Lock Manager is used to ensure consistency. Figure 4 shows Service 1 running on 3 PEs, and Service 2 running on 2 PEs. When a client connects to a database service, the CE uses a load balancing algorithm to decide which of the PEs offering that service should take the connection. Once the connection is made it remains until the client disconnects; during that time it services all transactions originating from that client. Where intra-query parallelism is exploited, slave processes in each PE are used to distribute and manage the execution of the query across a set of PEs.

Figure 5 shows the Distributed Access mapping of database servers onto GOLDRUSH.

In this mapping, again, each PE runs its own Database Server and the database tables are horizontally fragmented across PEs. As in the shared access mapping, a client connects to a PE and sends queries to it. However, each query is decomposed into sub-queries which are distributed such that each PE only has to access the table fragments which are stored on its local disks. The results of these sub-queries are then combined and the result returned to the client. The main effect of this type of mapping is that there is no use of the Distributed Lock Manager: only one PE (the local PE) can access a particular table fragment and so local locking can be used. However, if a query requires table fragments on more than one PE to be updated then some form of co-ordination such as Two Phase Commit (shown as 2PC in figure 5) is required.

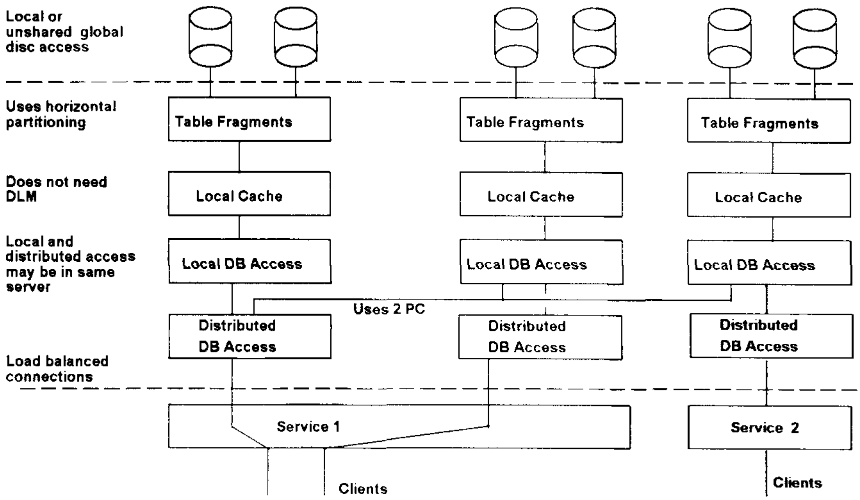


Figure 5 Distributed Access Mapping of Databases onto GOLDRUSH

In both the mappings, it is possible for both OLTP and Complex queries to be running against the same database simultaneously. In order to reduce the performance effects of one on the other, different sets of elements can be used for the two different types of queries – figures 4 and 5 both show two different services (Service 1 and Service 2) running on separate sets of elements but against the same database. It is also important that database tables are fragmented across sufficient numbers of disks to ensure that the IO throughput requirements of each type of query are met.

6. High Availability

The GOLDRUSH MegaSERVER is constructed from tens of commodity components such as processors and disks. Machines constructed in this way do not inherently offer the levels of reliability required to meet customer's requirements: in a simple parallel machine design, the failure of any one component may bring down the whole machine and halt the customer's application. However, careful design can exploit the parallelism to produce a system in which if one component fails then another takes its place. This may reduce the performance once the failure has occurred, but it will not halt the customer's application.

In this section we discuss how the parallelism of GOLDRUSH is exploited to provide high-availability. The main areas of resilience are:

6.1 Filesystems

For high availability, all database data is mirrored; the Veritas Volume Manager VxVM is used for this. Recovery from disk failure is totally transparent to the database servers accessing the data: when a disk fails, processing continues from the surviving mirror; the data is automatically re-mirrored onto one of the spare disks which are kept in the system for this purpose, and sometime later the failed disk can be replaced (using "hot pull and push").

In order to be resilient to element failures, each mirror is placed on a disk connected to a different element. Therefore, although when an element fails the disks locally connected to it become unreachable, all the data on those disks is still available on their mirrors (because they are connected to other elements).

The use of Veritas VxFS as the UNIX filesystem allows a filesystem which was owned by a failed PE to be reconstituted very quickly on another PE (by considerably reducing the time needed to restore filesystem consistency).

6.2 Processing Elements

Each element in the system runs its own instance of UNIX. This minimises the dependencies between elements and ensures that if an Element does fail then others can continue. Any parts of the Global Coherent Filestore which were owned by that element are logically moved to be owned by another element as described above. The database transactions being run on the failed element are automatically recovered by the database software running on another element using information found in tables, logs and journals. Any clients connected to a failing PE have to reconnect to the GOLDRUSH server. Those connected to other PEs see a short delay (up to 2 seconds) while the filestore is re-organised, but processing then continues.

6.3 Communications Elements

If a CE (or FDDI coupler) fails, then SQL connections from a client into GOLDRUSH will be routed automatically through another CE (or Coupler), provided that there is another route from GOLDRUSH to the client (and that the Client provides a full implementation of the transport connection protocols). This is achieved by maintaining configuration tables holding information on external sub-networks. Internal polling ensures that the failure of a CE or coupler is detected quickly. Following this, information in the configuration files is used to cause messages to be re-routed through another coupler, if this offers an alternative route to the Client.

6.4 Fans and Power Supplies

Spare fan and power supply capacity is present in the GOLDRUSH cabinets so that if one fails then the system can continue running. Uninterruptable Power Supplies can be used to prevent problems due to electricity supply failure.

6.5 Distributed Lock Manager

All lock information is mirrored in more than one PE so that the system is unaffected by element or process failure.

7. System Management

System Management is a vitally important, but often overlooked, attribute required by commercial parallel processors. Experience in the computer industry over recent years suggests that companies which downsize from a mainframe to a number of conventional UNIX systems often find that they make no overall savings in their IT budgets because even though the raw hardware and software costs may be lower, the cost of management increases dramatically. This is due to the fact that management tools for UNIX systems are generally more primitive than those available on mainframes, and because there is a need to manage a number of systems rather than just one. The danger therefore is that a parallel processor containing many UNIX Elements will prove to be extremely difficult and therefore costly to manage.

GOLDRUSH management tools have been designed to allow it to be managed as a single system, rather than as a collection of many UNIX systems. The basic architecture of System Management is shown in figure 6.

Each element runs an agent which offers local system management functionality, for example running commands and collecting statistics. The Management Element (ME) contains a layer of software which distributes management requests to the agents on the elements. The agents return results which are filtered and aggregated. Therefore the Management Element can offer to Management applications a single, high level interface for managing elements. The applications themselves run on the ME but are controlled from a PC (the System Management Workstation) or an external, possibly Enterprise-wide, Management Server. The Management Applications are designed to offer comprehensive coverage of the key management functions required by a database server.

A key concept in the management of GOLDRUSH is the use of named sets of components. Users can define sets of elements and then use the names of these sets in the system management applications, for example to monitor and administer the components of the set. Similarly, sets of disks and volumes can be defined and managed. The concept of sets is also key to resilience and tuning: if one element in a set running a database service fails another can be added to the set automatically. Because the management applications refer to the name of the set, and not the elements in it, then the change in the set contents is isolated from the manager.

The key management applications designed to provide single-image management of the parallel machine are:

7.1 Operations Management

This allows an operator to monitor the current status of all the major hardware and software components (the *managed objects*) of the system through a single pictorial representation. Each managed object is represented as an Icon whose colour represents the current status, and so changes of colour alert the operator of events in the system. Each managed object also has a set of actions associated with it, so allowing the operator to control it. For example, when a database service is created, a managed object is automatically created for it, and the status of the service (starting, running, in error, stopping...) is represented by its Icon's colour. The object also has a set of actions associated with it to allow the service to be started, stopped etc.

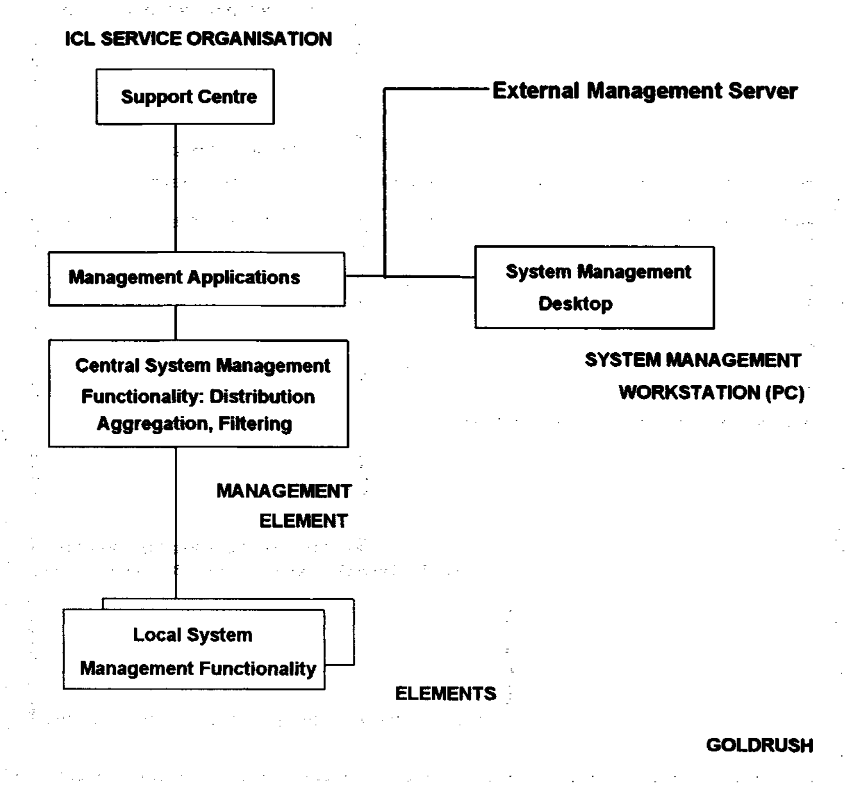


Figure 6 GOLDRUSH Management Architecture

7.2 Capacity Management

All the major hardware and software components can be monitored through this application, both in real time and historically. These components include disks, processors, the DLM, filesystems, database servers. The advantage of providing comprehensive monitoring of all levels of the system through a single interface is that it makes it possible to correlate related performance measures, such as transactions per second from the database server and processor utilisation from the kernel on which it is running. This aids both performance problem identification, tuning and trend analysis.

7.3 Configuration Management

This maintains the database of sets (described above), and provides interfaces through which they can be accessed. It also provides a graphical interface for configuring filestore. In a machine such as GOLDRUSH which may contain hundreds of disks it is not feasible for the system administrator to configure them one at a time, as in standard UNIX systems. This is particularly true when the complexity of configuring the mirroring is added. Therefore in order to simplify the configuration, the system manager can set up the filestore of one Element and have it replicated automatically across a set of Elements. This replication will take into account automatically the need to have disk mirrors held on separate PEs (see section 6).

7.4 Problem Management

Information on all problems observed within the system is passed to the Management Element where it is filtered and logged in a customer accessible database. If necessary they can be passed over a modem connected to the ME to the ICL service centre for action. This information may include problem evidence such as dumps. In the case of software faults, fixes can be passed back to the customer site.

7.5 Administration

This allows commands to be run on sets of elements, and provides a tool for installing software packages on a set of elements.

8. Conclusions

This paper has described the design of a commercial parallel computer system for running business critical database applications. The requirements for such systems are evolving as customers begin to take advantage of the new machines now available in the marketplace. We envisage a positive feedback effect in which these machines will open up new opportunities to users, who will then exploit them and in so doing place new requirements on the machines themselves. In particular, we expect that customers' realisations of the business advantages they can gain through the use of complex queries to analyse their corporate information will lead to demands for increasing performance. Similarly, the moves towards multi-media databases, and the advantages of centralising information in these servers, so making it all available for analysis, will increase the data capacity requirements.

GOLDRUSH is designed to meet these evolving requirements due to its scaleable parallel architecture, and its utilisation of commodity components whose capacity and performance are continually improving.

Acknowledgements

The GOLDRUSH MegaSERVER is the result of the work of a large number of people at ICL High Performance Technology, West Gorton, Manchester. We would also like to acknowledge our debt to our partners in the Alvey Flagship and ESPRIT EDS projects.

References

ALAMASI, G.S. and GOTLIEB, A.J., *Highly Parallel Computing*, Benjamin/Cummings, ISBN 0-80530-443-6, 1989.

HILDITCH, A.S. and THOMSON, C.M., *Distributed Deadlock Detection: Algorithms and Proofs* Dept. of Computer Science, University of Manchester, Technical Report Series UMCS-89-6-1, 1989.

INTERNATIONAL DATA CORPORATION, Source: International Data Corporation, 1992.

WATSON, I, WOODS, J.V., WATSON, P., BANACH, R., GREENBERG, M. and SARGEANT J., *Flagship: A Parallel Architecture for Declarative Programming*,. in Proceedings of the 15th Annual International Symposium on Computer Architecture, Honolulu, Hawaii, 1988.

WATSON, P., and TOWNSEND, P., *The EDS Parallel Relational Database System*, in Parallel Database Systems, ed. P. America, Lecture Notes in Computer Science 503, Springer-Verlag, 1991.

Biographies

Paul Watson

Paul Watson received a BSc in Computer Engineering from the University of Manchester in 1983, and gained a PhD from the same institution in 1986. From 1986 to 1989 he was a Lecturer in Computer Science at the University of Manchester but in 1990 he joined ICL, leaving in 1995 to take up a post in the Computing Science Department at the University of Newcastle.

During his time in Academia and Industry he has worked on the design of 3 generations of distributed store parallel systems. The first, funded by the Alvey Flagship project, resulted in the production of a prototype machine executing declarative languages. The second, the ESPRIT funded EDS machine, ran parallel declarative, language translation and database systems. The third is the ICL GOLDRUSH machine described in this paper.

Dr. Watson is a Member of the British Computer Society and is a Chartered Engineer.

George Catlow

George Catlow is the manager of the Parallel Systems Team within High Performance Technology based at West Gorton. He has many years experience in the development of computer related hardware and software, including robotics systems for automated manufacture, automatic test facilities for ICL's large Mainframes. He is a Chartered Engineer and Member of the Institute of Electrical Engineers.

The Hardware Architecture of the ICL GOLDRUSH MegaSERVER

Paul Watson[1] and Ted Robinson

ICL, High Performance Technology,
West Gorton, Manchester M12 5DR, UK

Abstract

This paper describes the hardware architecture of the ICL GOLDRUSH MegaSERVER, a distributed store parallel machine designed to operate as an open database server. The system is sold by ICL to commercial customers who require a high performance, high availability platform to support their critical business applications.

GOLDRUSH consists of up to 64 Elements, each of which can co-operate in database processing, exploiting both the parallelism found within complex queries and that found between queries in On-Line Transaction Processing workloads. The Elements are connected by a very high performance (1.2 GBytes/sec) internal network. Each Element has local disks to store the database. Some Elements have external comms connections allowing clients to connect to the system and so send queries into it for processing. Large tape libraries can be connected to multiple Elements, so allowing databases to be archived (and restored) in parallel at high speeds.

This paper discusses the architectural options for the design of high performance computers, explaining why the Distributed Store Architecture was chosen for GOLDRUSH. It then describes the design of the key aspects of the system including the parallel Elements, the internal network and the cabinets.

1. Introduction

The ICL GOLDRUSH MegaSERVER [Watson & Catlow, 1995] is an open, parallel database server designed for commercial users requiring high performance and availability. It has a distributed store, parallel architecture

[1] Current Address: Department of Computing Sciences, University of Newcastle, Newcastle-upon-Tyne, NE1 7RU, UK.

with up to 64 Elements, all of which can work together to exploit the parallelism found in database workloads. This includes both the parallelism found within complex queries and that found between queries in On-Line Transaction Processing workloads. Each Element has local disks on which the database data are held. The Elements are interconnected by a high performance Delta Network, while some Elements also have external comms connections allowing connection, via LANs and WANs, to clients which send queries into the parallel machine for processing. Large tape libraries can be connected to multiple Elements, so allowing databases to be archived (and restored) in parallel at high speeds.

The GOLDRUSH system evolved from earlier work carried out in collaborative projects with Universities and other companies. The two most important of these projects were: the Alvey Flagship project [Watson, et al, 1988] which produced a distributed store parallel system running both declarative languages and databases; and the ESPRIT EDS project [Watson & Townsend, 1991] which developed a parallel system for parallel databases, language translation and declarative languages.

This paper discusses the architectural options for the design of high performance computers, explaining why the Distributed Store Architecture was chosen for GOLDRUSH. It then describes the design of the key aspects of the system including the parallel Elements, the internal network and the cabinets.

2. Why choose a Parallel Architecture for Commercial Computing?

The main reason for considering parallel machine architectures to support database workloads is their ability to provide very high performance. While a single Element of a parallel machine may not exceed the performance of a high end Uniprocessor, the ability to harness sets of Elements to work together on the same problem makes it possible to construct systems whose performance is considerably greater, by a factor of 10 or more, than that of a Uniprocessor. A parallel machine also offers the attractive property of performance scalability. Over the lifetime of a customer's use of an application, the performance requirements usually increase. With a Uniprocessor, once the limit of CPU power has been reached, it is necessary to buy a new, more powerful machine (if one exists) to provide the extra power needed. In contrast, the power of a parallel architecture can be increased by adding Elements. This removes the need to buy a new machine, so reducing cost and minimising the disruption to users.

Another advantage of parallel architectures is that they are very cost effective. Figure 1 shows a graph of cost vs. performance for Uniprocessors. Initially the cost rises approximately linearly with the performance, but then the cost curve rises much more steeply and increases in performance become very expensive. This is the region where standard, relatively inexpensive commodity components are no longer fast enough to meet the required performance, and computer designers are forced to adopt more expensive solutions, including large static RAM caches with very low access times, interleaved stores, and leading edge, high speed semiconductor technology such as ECL.

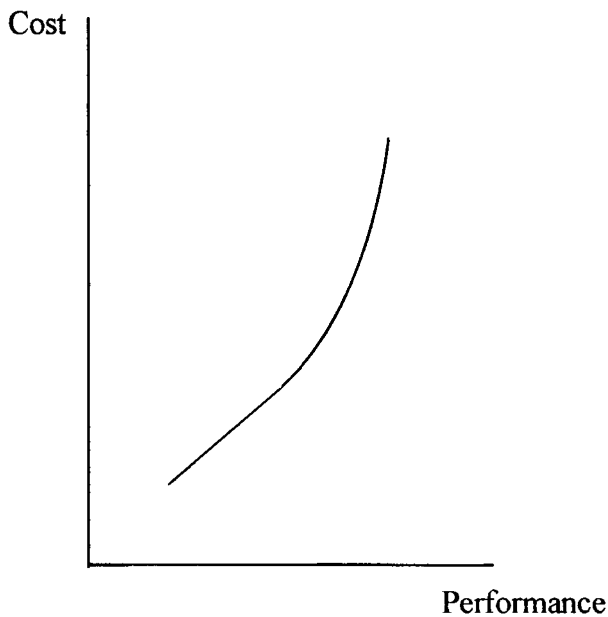


Figure 1 Cost vs. Performance for a Uniprocessor

Contrast this with the Cost vs. Performance Graph for a Parallel Machine (figure 2). Because performance can be increased by adding Elements, rather than by increasing the power of the Elements, the cost vs. performance graph rises in a much more linear fashion. The steps in the graph reflect the fact that due to the need to package Elements into cabinets, there will be points where adding an Element also requires the addition of a new cabinet.

One of the key issues in parallel machine design is choosing the right performance for the Elements. Given the cost vs. performance graph for a Uniprocessor shown in figure 1, it is sensible to aim for an Element performance on the knee of that curve. Choosing a performance below the knee reduces Element cost, but more Elements are needed to achieve a particular performance, and this increases the cabinet & network costs associated with those Elements. It also increases the amount of parallelism needed in the software to achieve a particular level of performance. Above the knee of the curve, increases in Element performance require relatively large increases in cost, and it is cheaper to achieve these performance gains by adding more Elements to the machine.

Another advantage of choosing an Element performance on the knee of the curve is that it makes it possible to use commodity products in the Element design. Not only does this reduce cost but it also reduces development time and so decreases time to market when compared with the alternative of building a very high performance Element from low-volume or specially designed components. It is also an advantage that using commodity products provides a straightforward, low-cost path for future performance upgrades. For example, the producers of the major microprocessor families are continually releasing higher performance but compatible versions of their products. This provides the opportunity to increase the Element performance by adopting the faster version of these components as they become available. It is important that the Element is designed to allow this by ensuring, for example, that other parts of the design such as the bus and internal network connections do not become bottlenecks if the performance of the microprocessor increases.

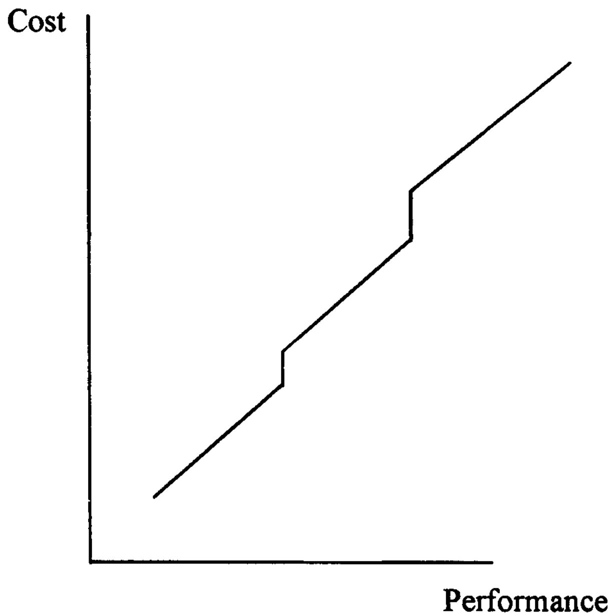


Figure 2 Cost vs. Performance for a Parallel Machine

Of course, the potential for high performance provided by the raw hardware of parallel machines is only realisable if the system and application software can exploit it. This has been shown to be true for database servers where there is significant parallelism between transactions in On-Line Transaction Processing workloads, and within complex queries [Watson & Catlow, 1995].

3. Classes of Parallel Architecture

There are two basic types of parallel architecture to consider for commercial systems: Shared Store Multiprocessors and Distributed Store Processors (often known as MPP – Massively Parallel Processors). These are considered in this section along with the explanation for the choice of the architecture for GOLDRUSH.

3.1 Shared Store Multiprocessors

Figure 3 shows the basic architecture of a Shared Store Multiprocessor. This consists of a set of processors communicating with store and IO (Disks and Communications Devices) over a shared bus. This architecture is attractive because all processors have access to the same store, which simplifies writing parallel programs. However, there are two potential bottlenecks which limit the scalability: the bus and the store. All processors share one store, and so it must be fast enough to service all their requests. This tends to become a problem when more than a small number of processors are sharing store. There are (expensive) solutions such as store interleaving [Stone, 1993] which alleviate this problem, but this still leaves the second bottleneck: the bus. Each unit in a Shared Store Multiprocessor communicates with the other units over the bus. Therefore as processors are added to increase performance, the bus becomes more heavily used. However, because it has a fixed bandwidth, the share of the bus available to each processor decreases, which limits scalability. This problem has been getting worse over time, as the processors themselves become faster, and so require greater bus bandwidth. However, it is not possible to produce similar performance increases in the buses themselves because they are basically passive tracks of copper on a printed circuit board, with significant reactive load, whose speed is limited by the time it takes for changes in signals to propagate down the tracks. For this reason, at present buses are limited to around 70MHz.

The problems of bus and store bottlenecks have limited the number of processors typically to 16 in Shared Store Multiprocessors. However, due to increases in processor performance it is expected that the next generation of these machines will have more limited scalability: probably 8x or less. This therefore reduces their performance gains over Uniprocessors.

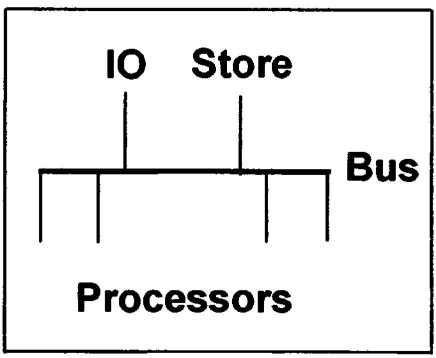


Figure 3 The Shared Store Multiprocessor Architecture

3.2 Distributed Store Multiprocessors

Figure 4 shows the Distributed Store Multiprocessor architecture. This consists of a set of Shared Store Multiprocessor Elements connected by a high performance network. The key constraint of this type of architecture is that each processor can access only its local store, and so all communication between Elements is via explicit messages sent over the internal network. Considerable research work is being undertaken in both academia and industry to develop mechanisms to provide the "illusion" that all store is equally accessible from all processors, but the main problem with this is the large difference in latency between a local store access and a remote access: the local cache hit rate must be very high if remote accesses are not to reduce performance significantly.

The key advantage that the Distributed Store architecture has over both Uniprocessors and Shared Store Multiprocessors is its scalability. By adding Elements, all the key performance attributes are scaled including: processing, bus, store, IO bandwidth and connectivity. This scalability is only realisable if the network interconnecting the Elements does not become a bottleneck. However it is possible to achieve this using today's technology (see section 7 for a description of the GOLDRUSH internal network). In fact, there are no architectural limits to the scalability of the Distributed Store hardware, though practical limits may be imposed by limitations in the scalability of the software or the cost of very large machines. For these reasons, most Distributed Store systems are limited to tens rather than hundreds of Elements. However, this still gives them a significant performance advantage over Shared Store Multiprocessor machines, and an enormous advantage over Uniprocessors.

To the system programmer writing a parallel application, the Distributed Store architecture is more difficult than the Shared Store Multiprocessor and Uniprocessor as it is necessary to sub-divide the program into units which run on different PEs and communicate via messages sent over the network. How this is achieved for the GOLDRUSH system is described in [Watson & Catlow, 1995].

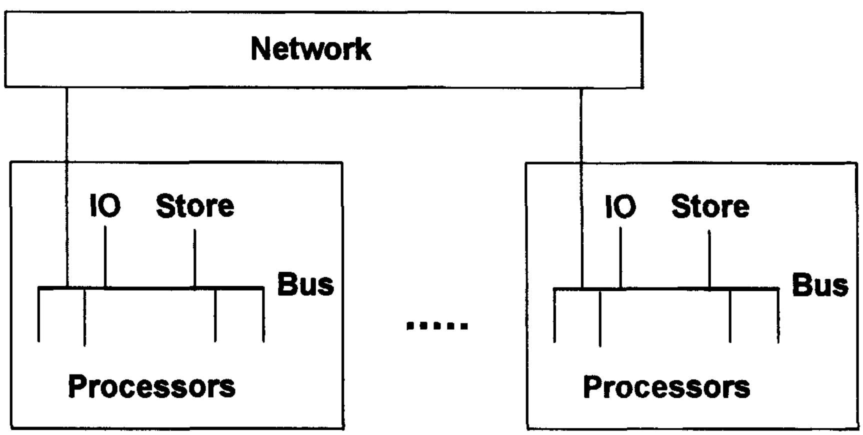


Figure 4 The Distributed Store Multiprocessor Architecture

Because of the scalability advantages of this architecture, resulting in a higher overall performance, this is the architecture adopted for GOLDRUSH.

4. Hardware Architecture Overview

The hardware architecture of a GOLDRUSH system is shown in figure 5. It consists of a set of up to 64 Processing Elements (PEs), Communications Elements (CEs) and Management Element (ME) connected together by a high performance network (DeltaNet).

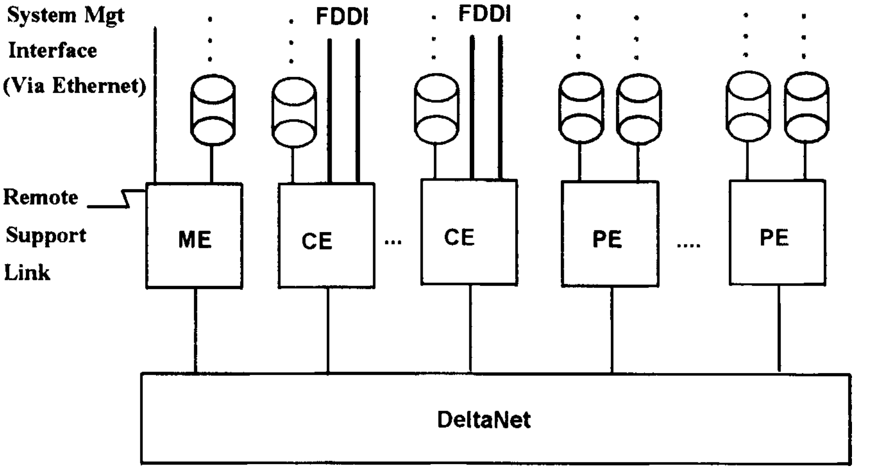


Figure 5 The Internal GOLDRUSH Architecture

The PE is the most common Element and is designed to run a database back-end. It has a connection to the internal DeltaNet, and up to 12 disks can be locally connected.

The Communications Element is identical to the Processing Element except that it also has two FDDI couplers for client connection. Multiple CEs can be configured in a system to improve both performance and resilience.

The Management Element is a conventional mid-range UNIX processor (currently an ICL DRS6000) which runs the Management Software. Its responsibilities include controlling the establishment of the system, and problem diagnosis. It contains a "Teleservice" modem connection which allows problem reports to be sent to a service desk, and remote problem diagnosis.

For high performance archiving, tape libraries can be attached to the Elements via their SCSI connectors. Connections from multiple Elements to the tape library provide parallel archiving (and restoration) by allowing multiple data streams to be archived (and restored) simultaneously to multiple tape drives.

The following sections describe the design of the main components of the GOLDRUSH Hardware: the Processing Elements (section 5), the Communications Elements (section 6), the DeltaNet (section 7), the Cabinets (section 8).

5. Processing Elements

The Processing Element (PE) is the basic building block of the GOLDRUSH machine. Its job is to run a UNIX operating system and Database Server efficiently. The main requirements it has to meet are therefore:

- high processing power
- high performance connection to the DeltaNet
- large main store to hold the database cache
- high disk connectivity so that large databases can be stored

It was decided to base the PE around SPARC processors because of their high performance and the availability of support chips. A standard SPARC Mbus is also used to connect the components within a PE. Figure 6 shows a schematic diagram of the PE hardware.

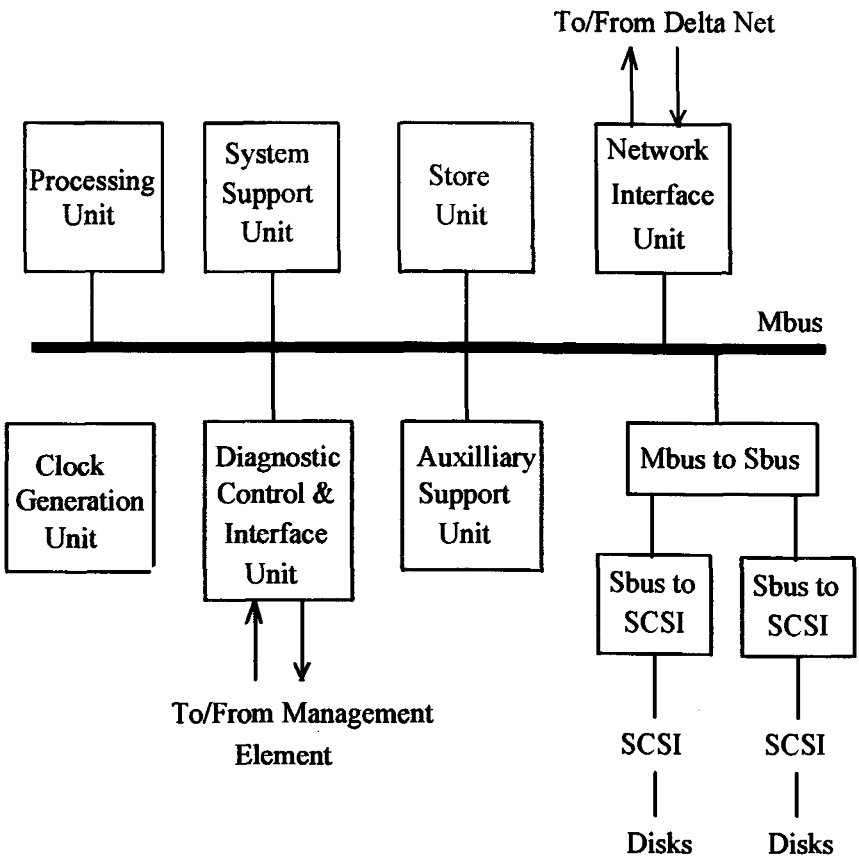


Figure 6 The Processing Element Design

The main units in the PE are:

1) Mbus: the standard SPARC processor bus, which operates at 40MHz and is 64 data bits wide. The Level 2 industry standard protocol is supported.
2) Processing Unit: a SPARC processor which runs the operating system and application software. Currently the SPARCs used for the Processing Unit (and the System Support Unit described below) are 90MHz parts but as new, faster SPARCs are produced by chip manufacturers, these can be used.
3) System Support Unit: another SPARC processor dedicated to sending and receiving messages through the DeltaNet, so reducing the load on the Processing Unit and improving system performance. When a process on the Processing Unit wishes to send a message to another PE, it calls a system interface which passes the start address and length of the message to the System Support Unit. This is then responsible for fragmenting the message into packets which can be sent over the DeltaNet, and ensuring that they arrive without error. Similarly, when a message is received by a PE, it is the System Support Unit which de-fragments it (having waited until all the DeltaNet packets used to transmit the message have arrived), checks that there has been no transmission error, and places the message in store where it can be read by the receiving process. Only then is the Processing Unit informed of the arrival of the message.
4) Store Unit: a large (currently 256MByte) main store built from (currently 16MBit) Dynamic RAMs. The data width is 64 bits, with a further 8 bits for Hamming error correction and detection. For high performance, the store is 2-way interleaved.
5) Network Interface Unit: this serves as a DMA unit between the Store and the DeltaNet. When a packet is to be sent, the System Support Unit provides the Network Interface Unit with the physical store address and length of the data. The interface to the DeltaNet operates asynchronously with the Mbus at a frequency of 20MHz. Two Cyclic Redundancy Code (CRC) bytes are appended to the message on transmission. On receipt of a message these bytes are checked. Additionally, the length of the packet is checked against that specified in the packet header. Also checked is the PE destination address in the header to ensure that the packet has not been misrouted.
6) Clock Generation Unit: generates 40MHz Mbus clocks to all Units.
7) Auxiliary Support Unit: provides a collection of miscellaneous functions including timer, interrupt generator and bootstrap EPROM.
8) Diagnostic Control and Interface Unit: The GOLDRUSH Management Element (ME) has a diagnostic RS-232 connection to each Element, and to the DeltaNet. This is used to co-ordinate the establishment of the machine by sending boot commands to each Element, but is used also to allow the ME to access diagnostic information if an Element cannot be reached over the DeltaNet because it has failed in such a way that it can no longer send or receive messages. The Diagnostic Control and Interface Unit provides this connection from the PE to the ME.

9) Connection to Local Disk is provided by two fast and wide SCSI-2 connections, each of which can connect to up to 6 Disks. They have a peak transfer rate of 20MBytes per second. The SCSIs are connected to the Mbus via Mbus-Sbus and Sbus-SCSI bridges. The Disks currently used are standard 3.5" diameter devices with a capacity of 4GBytes. Because the disks are of standard size and interface, this specification will be upgraded as new, higher capacity devices become available from manufacturers.

The physical dimensions of the PE are 364mm x 233mm x 41mm.

6. Communication Elements

The Communication Element (CE) is exactly the same as the Processing Element except that one of the Sbus-SCSI bridges is replaced by two Sbus-FDDI (Fibre Distributed Data Interface [Taylor, 1992]) bridges. This provides each CE with two optical FDDI couplers. FDDI was chosen as the external comms interconnect because of its high performance (100 Mbits per second) and because bridges are available from it to virtually all other types of LAN and WAN, including Ethernet.

In the GOLDRUSH system, the CEs do not run database servers, but instead are dedicated to relaying messages transparently between the PEs and the FDDI couplers. This is done by software in such a way that each Processing Element behaves as if it had local FDDI connections.

The number of CEs per GOLDRUSH machine is variable, but 1-2 per 16 Elements is typical. The exact number is determined by the need to provide sufficient external comms bandwidth for database queries, and also resilience to CE failure: if a CE fails then comms messages will be re-routed through another CE provided that there is a route from it to the external client.

7. DeltaNet

The internal network is a key component of any Distributed Store Parallel Machine because if it forms a bottleneck, then it will prevent the system's performance scaling as Elements are added. Also, if its latency is too high, then this can reduce the response time for database queries to unacceptable levels. The DeltaNet is therefore designed to provide both high throughput and low latency. The design was pioneered in the Flagship machine and then enhanced in the EDS project.

The DeltaNet is a multi-staged network built from the basic building block of an 8 x 8 crossbar switch (called a Network Switching Element: NSE) which provides unidirectional channels dynamically from each of its 8 inputs to any of its 8 output ports. These NSEs are connected in stages: two stages allowing up to 64 connections. Figure 7 shows how a 32 Element network can be created from these switches. Note that the unconnected NSE inputs and outputs would be used in a 64 Element machine. They can also be used as alternative paths in systems of up to 32 elements.

Each packet transmitted through the network consists of 128 Bytes of data, 16 Bytes of header information and a 2 Byte Cyclic Redundancy Check. The header includes the destination Element, the source Element, the length, and a sequence number which is used if a message of greater than 128 Bytes has to be fragmented (by the System Support Unit) and sent in a set of 128 Byte packets. The 2 Byte Cyclic Redundancy Check is generated by the sending Element, and checked by each NSE the packet passes through – so ensuring the integrity of the transmission. Connections from the Elements to the NSEs are 11 bits wide: 8 data plus 3 control.

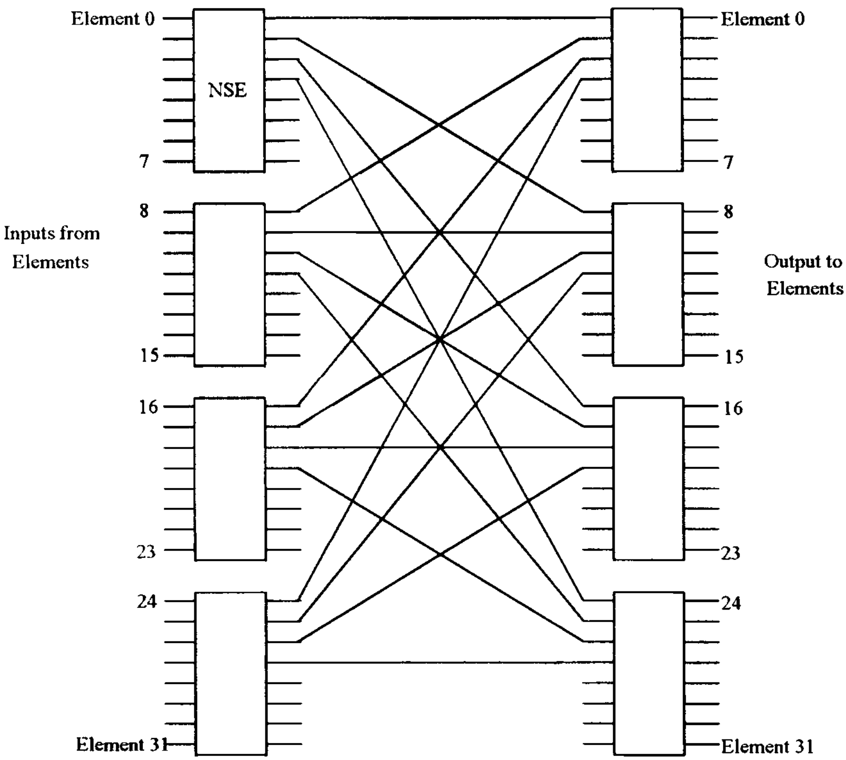


Figure 7 A 32 Element DeltaNet[2]

When a packet is sent by an Element it reaches the first NSE where the Destination address is inspected and the packet is immediately routed out along the appropriate path towards its destination. When it reaches the second stage of the network, this process is repeated and this causes the packet to be routed to the destination Element. Therefore, if there is no contention, a packet will pass through both Network stages without any buffering, and this

[2] The cross connections not shown follow the same pattern as those indicated; they are omitted from the diagram purely for clarity.

reduces the latency of the network. However, if two or more packets arrive at the inputs of an NSE simultaneously, and each needs to be transmitted through the same output, then this results in contention as only one packet can be transmitted through an output at any one time. The Network deals with this as follows. Each NSE has 4 buffers on each input. If a packet arrives at an input, and its output is blocked then it is held temporarily in one of these buffers until the output is free. In the worst case, if all four buffers at an input are full, then no more packets are accepted on that input until a buffer becomes free.

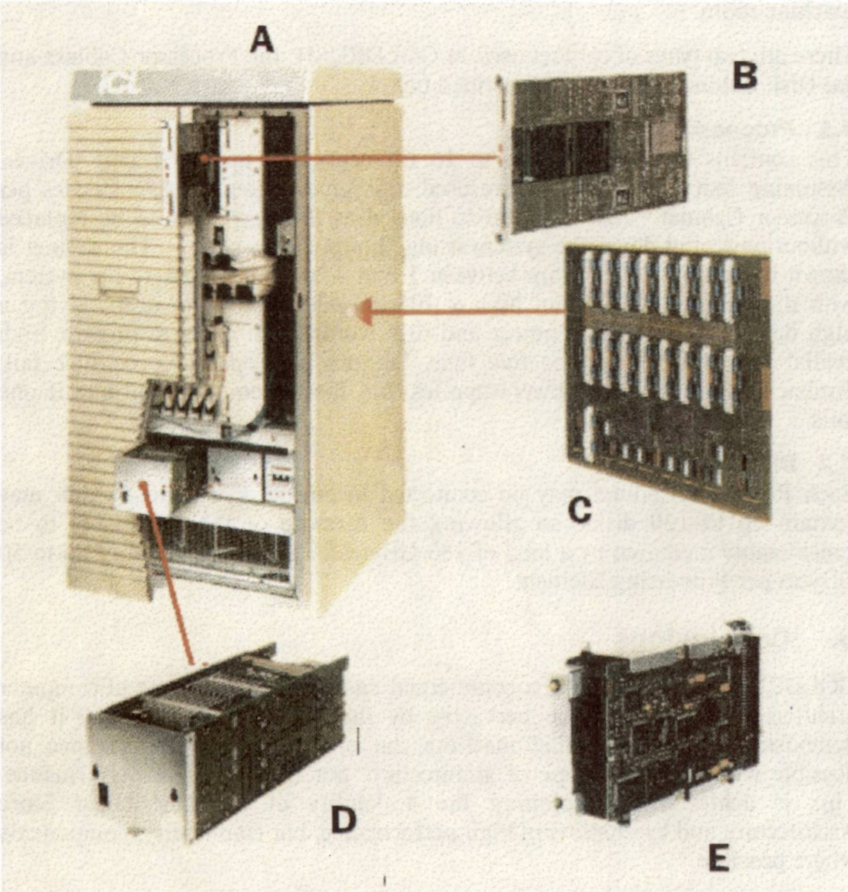


Figure 8 The GOLDRUSH Processor Cabinet shows the front view of the cabinet A, and some of the internal units; B is a processor element, C is a backplane providing DeltaNet functionality, D a power supply and E is a disk unit.

The Delta Network runs asynchronously with respect to the Elements, currently at 20MHz. In combination with the System Support Unit, it allows each Element to send and receive data at up to 20MBytes/s simultaneously. For a 64 Element system, this provides a total bandwidth of 1.2 GBytes/sec.

8. Cabinets

The design of the cabinets for a parallel machine is important because of the need to support a wide variety of configurations – GOLDRUSH machines can have from 4 to 64 Elements – and the importance of limiting the space occupied by the larger configurations so they do not require a very large machine room.

There are two types of cabinet used in GOLDRUSH: the Processor Cabinet and the Disk Cabinet. These are described below.

8.1 Processor Cabinet

This contains the DeltaNet, up to 16 Elements and up to 64 Disk Drives. Assuming that 4 GByte disks are used, this gives a total of 256 GBytes per Processor Cabinet. All disks are so housed as to allow them to be replaced without powering down the system using "hot pull and push". The cabinet is shown in figure 8. There are between 1 and 4 of these Cabinets per system, with dimensions of 1450mm high x 705mm wide x 860mm deep, giving a high density of processing power and disk storage. In order to provide high availability, the cabinet has five fans, but can still operate even if 2 fail; Similarly, there are two power supplies, but the system can continue if one fails.

8.2 Disk Cabinet

Each Processor Cabinet may be connected to one Disk Cabinet, which may contain up to 120 disks, so allowing the filestore of the Processors to be considerably increased by a total of 480 GBytes. This gives a total of up to 50 GBytes per Processing Element.

9. Conclusions

GOLDRUSH is one of the first commercial embodiments of a type of computer architecture expected to be pervasive by the end of the century. It has demonstrated how a parallel machine can offer levels of performance not possible with any other type of architecture, but with low cost-performance. This is achieved by exploiting the scalability of the Distributed Store Architecture and by employing high performance, but commodity, components where possible.

Acknowledgements

The GOLDRUSH MegaSERVER is the result of the work of a large number of people at ICL, High Performance Technology, West Gorton, Manchester. We would particularly like to recognise the contribution of all past and present members of the GOLDRUSH Hardware Team. A debt is also owed to our partners in the Alvey Flagship and ESPRIT EDS projects.

References

ALAMASI, G.S. and GOTLIEB, A.J. *Highly Parallel Computing*, Benjamin/Cummings, ISBN 0-80530-443-6, 1989.

STONE, H.S., *Highly-Parallel Computer Architecture*, Addison Wesley, ISBN 0-20152-688-3, 1993.

TAYLOR, M. FDDI – The High Speed Network for the Nineties, *ICL Tech.J.* 8(2) pp. 225-241, 1992.

WATSON, P. and CATLOW, G.W. *The Architecture of the ICL GOLDRUSH MegaSERVER, Ingenuity.*10(2), pp. 191-205, 1995

WATSON, I, WOODS, J.V., WATSON, P., BANACH, R., GREENBERG, M. and SARGEANT, J. *Flagship: A Parallel Architecture for Declarative Programming,*. in Proceedings of the 15th Annual International Symposium on Computer Architecture, Honolulu, Hawaii, 1988.

WATSON, P., and TOWNSEND, P. *The EDS Parallel Relational Database System*, in Parallel Database Systems, ed. P. America, Lecture Notes in Computer Science 503, Springer-Verlag, 1991.

Biographies

Paul Watson

Paul Watson received a BSc in Computer Engineering from the University of Manchester in 1983, and gained a PhD from the same institution in 1986. From 1986 to 1989 he was a Lecturer in Computer Science at the University of Manchester but in 1990 he joined ICL, leaving in 1995 to take up a post in the Computing Science Department at the University of Newcastle.

During his time in Academia and Industry he has worked on the design of 3 generations of distributed store parallel systems. The first, funded by the Alvey Flagship project, resulted in the production of a prototype machine executing declarative languages. The second, the ESPRIT funded EDS machine, ran parallel declarative, language translation and database systems. The third is the ICL GOLDRUSH machine described in this paper.

Dr. Watson is a Member of the British Computer Society and a Chartered Engineer.

Ted Robinson

Ted Robinson obtained a BEng in Electrical Engineering from Liverpool University in 1958. After a period with AEI he joined ICL (then ICT) in 1967 as the designer of a large fixed disk controller for 1904E machines. Subsequently he has been involved in a variety of hardware projects – System 4 emulator in 2970; 2966 OCP design; and clock and store design for distributed system 39 (DM1).

For the past 9 years he has had numerous responsibilities in parallel architecture collaborations – Alice, Flagship and EDS, the latter evolving into the GOLDRUSH MegaSERVER. He is currently investigating options for GOLDRUSH enhancements and replacements.

CAL in Higher Education – Potential and Pitfalls

P. H. Ford

Dean of Science, University of Nottingham, UK

J. E. Graves

Network and Applications Support Manager, University of Manchester, UK

Abstract

The University of Nottingham, in common with many other institutions of Higher Education, sees the increased use of multimedia as a key element in future planning and development. The long-term goal is to bring about a culture change whereby academics of all disciplines regard skills in courseware development as a natural adjunct to their existing teaching tools, on a par with use of the blackboard, chalk and the overhead projector. The aim is, therefore, to embrace Computer-Assisted-Learning (CAL) as an integral part of the overall learning environment. The rationale for this, together with the key issues which need to be addressed in order to bring this about, are discussed in the context of ongoing developments at the University. In particular, the issues of an adequate IT infrastructure, choice of authoring tools, delivery strategies, standards, training and support are highlighted.

1. Introduction

Higher Education over the last ten years has seen radical changes. The distinction between Universities and Polytechnics has disappeared and student numbers continue to increase. From a ratio of around 1:10 of all eighteen year olds moving into Higher Education in the early 1980s, we now have a ratio of around 1:5, with the likelihood of further pressures to move to a ratio of 1:3 by the turn of the century.

These very substantial increases in the student population have inevitably caused great pressure on the infrastructure within each institution and have meant a considerable reduction, in cost terms, in the unit of resource available to each student. In parallel with this, there has been increased emphasis by Government and the Higher Education Funding Councils for greater accountability and quality management in both research and teaching.

Pressures on staff time have, therefore, focused attention on ways in which increased productivity can be obtained in the context of the student learning process without any reduction in quality and, indeed, where possible, leading to improvements to the overall learning experience of the student.

The above has to be viewed in the context of the important and fundamental changes which have taken place in the use of information technology within Higher Education over the same period. The majority of institutions have invested heavily in high speed campus networks, to satisfy their research and administrative requirements plus academic service functions such as Library automation. These networks are most often based on 100 megabit per second FDDI backbones with UTP wiring to the desk top. Distributed computing is now established practice with client-server systems based in departments, faculties or buildings as dictated by the relevant topology. Hierarchical distribution of servers for central functions, faculty use and departmental activities is now the norm, with a proliferation of public access networked workstation clusters distributed throughout the campus, allowing students to work at their own pace, from whatever location they consider most convenient at the time. There is thus a gradual shift of emphasis away from the traditional style of learning comprising formal lecture plus formal tutorial reinforced by private reading, towards the concept of the University being an integrated "total learning environment". In theory at least, the technology is in place to allow students to have networked access to a wide range of study material in addition to continued provision of personal or group tuition by academic staff.

2. CAL Development

At the end of the 1980s, the University of Nottingham took a strategic decision to explore the potential for widespread use of Computer-Assisted-Learning across all disciplines; taking into consideration the extensive IT provision already in place or being planned, and the increase in student numbers forecast within the coming decade. The intention, although recognised as a distant goal, was gradually to introduce a culture change in teaching methods whereby lecturers would become as comfortable developing interactive courseware for student use, as they already are with the traditional teaching tools of blackboard, overhead projector and video cassette. It was acknowledged that to realise this strategy a number of objectives must be met, namely; adequate access for all staff to development facilities; courseware development tools independent of traditional IT literacy; sufficient and appropriately configured workstations for student use; and availability of training and support. Furthermore, it would not be sufficient merely to promote the culture change but the effectiveness and quality of the resulting courseware should also be ensured.

3. Key Factors

To progress the concept and move towards these objectives, the essential considerations were felt to be:

3.1 The Choice and Potential of Courseware Development Tools

This was perhaps the most important factor of all and raised a number of issues. To minimise the impact on technical staff of the training and support requirements of the academic authors it was thought essential to promote a limited number of authoring tools for use by the community. It was recognised that the vast majority of the University's networked workstations comprised either IBM-compatible PCs with a variety of specifications or Apple Macintosh equipment, also of varying types. It was thus highly desirable that courseware development tools should provide both cross-platform development and delivery, wherever possible. Since appropriate use of animation, video and sound was envisaged, the tools should enable this to be achieved easily. To attain the aim of the vast majority of lecturers participating in courseware development it was essential that the use of chosen tools did not require extensive IT skills. Also, the spread of disciplines across the University and, in particular, the differing pedagogical requirements of the Humanities from Science and Engineering, suggested that more than one tool would be necessary. Discussions revealed that two broad styles of courseware should be considered; the hypermedia approach for browsing and exploring the information and the more directed approach for learning about and testing specific knowledge areas.

A survey of available tools suggested that the older tools would require programming skills beyond the reasonable aspirations of the average academic. However, the emergence of new tools with more intuitive graphical user interfaces reinforced the view that courseware could indeed be developed without extensive IT skills on the part of the developer. At the time, a number of tools were available and extensive evaluation yielded the conclusion that a combination of Authorware Professional and Guide would best meet the criteria described earlier. Authorware offered the potential for cross-platform development and delivery. Guide did not, but was at least available on both platforms and transition from one to the other was not felt to be technically demanding. The University recognised that, to encourage the use of these recommended tools, it would be beneficial if they were freely available to the potential authors and hence negotiated site licences for both products. This approach has proved highly successful in limiting the number of tools supported and by maximising skills transference between the academic staff.

Although other competing courseware tools are now available, both Authorware and Guide have been in widespread use within the University for the last four years and they continue to provide a sound basis for ongoing courseware development.

3.2 Courseware Delivery

Almost as important as the choice of tools was the courseware delivery strategy. The use of CD-ROM was considered inappropriate within the University environment for the delivery of courseware on the scale envisaged. Delivery across the network, however, offers the potential for courseware to be updated from a single place plus the possibility for authors to gather feedback, usage statistics and student test results as required. A discussion follows on the campus network and associated factors which have influenced the evolutionary delivery strategy adopted.

The network infrastructure comprises an FDDI backbone running at 100 megabits per second, with departmental ethernets attached to the backbone and supported on Unshielded Twisted Pair (UTP) wiring. Around 4,000 sockets are supported on the network, with attachments to IBM PCs, UNIX Workstations, Apple MacIntosh systems, as well as a wide range of departmental and central servers. At least in theory, this topology should allow a student at an appropriate workstation to be able to use interactively any available courseware mounted on a central courseware server. However, the reality of such a usage pattern soon highlights the deficiencies of such a network topology, posing severe traffic overloads on various network links.

An analysis of the pattern of student use shows that it is confined mainly to the academic department or building during the normal Monday to Friday teaching time. Even then, however, students not undertaking intensive teaching may choose to access courseware from any part of the network where clusters of publicly available workstations are positioned. During the evening and at weekends the pattern of usage, and thus the network traffic load, changes significantly. Access to courseware at these times is predominantly via the clusters of publicly available workstations or from halls of residence. The evening and weekend periods also see a potential demand from students who live off campus and who therefore feel disadvantaged if access to their courseware is not available to them in their home environment.

Solutions to these problems are still being evolved. The concept of one central server for all courseware cannot on its own support the demand without major upgrades to the speed of the network, and this is unrealistic at the present time in terms of cost. The strategy currently adopted is to maintain a complete set of courseware on a central server. This is not directly accessible by students, but departmental server managers can then copy appropriate elements onto their departmental servers. During the week the stability of the network throughout the normal working day is effected through suitable bridging. Demand during the evenings and weekends, although significant remains at a level such that access to departmental servers across the network is acceptable. The peaks in demand during pre-examination periods indicate that this is unlikely to be true in the next academic session as the volume of available courseware and consequent student demand increases.

The next phase in the strategy is to duplicate the most widely used courseware on the most appropriate local server(s). However, the final stage in this strategy is to monitor the requests for courseware and download automatically to the appropriate servers that most frequently requested at times when there is least traffic on the network. It is not anticipated that the interval between such distributions would be less than a day. For those students wishing to use courseware not in the popular core it will be necessary to provide a procedure such that they can request specific items to be included in the dynamic distribution. If the majority of the students can be satisfied from the core courseware selection, it should be possible to allow requests for less popular items on demand, giving students the flexibility of a central server with the least impact on the network.

To date, the needs of students in their own homes has not yet been addressed. However, the telephone exchange has been integrated with the campus network, thus ensuring that the network can, in the medium term, be accessed through ISDN. Work is now ongoing to provide an alternative delivery mechanism whereby students can dial up over the public switched network by modem or digital connection and download their courseware onto their local PC. Such a delivery mechanism should eliminate the need for high-speed network connection, but needs modern data compression techniques in order to transfer effectively the very extensive files involved. In addition, connection via the local cable company should provide higher speed access to major sections of the local community.

3.3 The Student Interface

As the volume of available courseware modules has expanded, both the identification and location of the courseware have become increasing problems. Currently each server offers its own menu of options including CAL. However, even at the level of departmental servers the number of courseware modules is such as to require more than one level of menu. As a result, attention is now being directed at an appropriate common user interface at the student workstation providing easy browsing facilities to navigate what is available and where. This concept is to be built into the overall menu scheme of all available campus network facilities, thus minimising the demotivating effect for the student of an extensive search procedure before study can commence.

3.4 Design Standards

It was acknowledged that the recommendation of standard, relatively easy-to-use tools was not in itself sufficient to promote the development of good quality and easy to use courseware. In particular, the interface presented to students should be as intuitive as possible and adhere broadly to standards of presentation such that use of courseware authored by a number of different academics would appear similar. Thus students could concentrate on the subject matter and not the interface. Accordingly, a project was undertaken by experienced developers to determine and document the most relevant standards [Davies & Brailsford, 1994]. A number of issues were identified including:

- ease of cross platform development/delivery
    - authoring features which should be avoided in order to allow courseware to be transferred easily to alternative platforms
- screen layout
    - avoidance of cluttered screen appearance by the logical organisation of objects
    - the requirement for visible connection of interrelated objects or concepts within the same learning space when using different media

- paging and scrolling
    - the preference for paging when presenting the main text
    - the appropriate use of scrolling where paging would appear artificial, for example when displaying long tables or historical texts
- readability
    - advice on the suitability of using upper or lower case text, justification, serif or sans serif fonts and font sizes
    - guidance on background and foreground colours and the use of emboldening, italics, underlining etc.
- visibility
    - positioning messages, displays and buttons for ease of use and understanding
- colour
    - recommendations on the use of colour including colour coding and colour combinations to avoid.

Authors are encouraged to use these guidelines, although it is recognised that originality and creativity should not be stifled by strict adherence to standards where it would obviously be inappropriate.

3.5 Training and Support

To promote the change in institutional culture, a training and support unit was established comprising advisors skilled in courseware development, initially working alongside a small core of enthusiastic academic staff to produce the early products. In addition, the specialist skills of a Graphics Designer were considered to be essential within the unit, together with provision of equipment and software to facilitate image generation and enhancement, animation, text and image scanning and photographic and video capture.

During the early stages of the initial pilot projects, monthly meetings were organised, each comprising an informative talk and demonstration followed by discussion. These meetings proved so useful that they continue to be held and are open to anyone interested in CAL and as such provide a beneficial inter-disciplinary forum for discussion of ideas. Alongside these meetings a number of promotional events have been held at both departmental and faculty level, with the aim of encouraging academic and service staff to consider how CAL might be used to support the learning environment. To assist new developers, the unit has also been active in developing reusable elements including shells, buttons and paging models together with standard modules, for example to provide a framework for multiple choice questions or to gather feedback or monitoring information. A database of graphic images is also being established. Where possible, developers are encouraged to use external files to remove the dependency of the content from the structure, enabling the content to be modified. However, before undertaking any development, potential authors are advised to consider; the purpose of proposed courseware, the scope of the information to be presented or tested, the learning strategy to be adopted, the relevance of incorporating different media and whether sufficient interaction is planned to stimulate users' interest: thus a significant amount of time is spent producing a storyboard before beginning to author.

This central unit has proved essential in the ongoing development of the project, and careful 'seeding' of projects around the campus is now bringing its own reward in terms of other staff seeking to emulate and surpass the achievements of the pioneers.

3.6 Top Management Support

To bring about a culture change on the scale proposed, it has been essential that senior management are fully supportive of the initiative. The University has been fortunate in having this support from the Vice-Chancellor and his senior colleagues, and the concept of increased and widespread use of CAL is encapsulated within the University's published strategic plan. An IT Strategy as a support document for this strategic planning is also in place, outlining appropriate workstation configurations, network developments, and provision of training and support.

4. National Initiatives in Courseware Development

Most of the above issues were identified, although by no means solved, during the late 1980s and early 1990s. A major boost, however, was provided to the initiative in 1992 by the announcement of the Teaching and Learning Technology Programme (TLTP) by the Higher Education Funding Councils of England, Scotland and Wales, together with the Department of Education in Northern Ireland. This initiative sought to invest substantial funding to initiate a number of courseware development projects. These projects were either on the basis of a consortium of departments of similar discipline across a number of Higher Education Institutions or were 'institutional' projects specifically designed to assist the culture change described earlier. By virtue of its early experience, Nottingham was well placed for many of its departments to bid to become a member of a consortium, but the University also obtained funding for an institutional project which substantially increased the staffing and hence the support available from the central support unit. This enabled each member of the unit's staff to be more closely identified with the CAL support of specific disciplines assigned to them, to the mutual benefit of both developer and support officer.

5. Evaluation

Courseware from the first year's priority projects has been delivered for one academic session only, so evaluation is still in process, and the feedback is preliminary. A standard feedback model, to attach to the end of courseware, has been developed so that users can express their feelings about the course. Additionally, statistics are being gathered centrally on the usage of the courseware; generally uptake has been good and the feedback has been positive. Even at this stage other encouragement has been forthcoming in the form of encouraging comments from external examiners. One of the main tasks for the current, and next, year is to attempt to evaluate fully the success of the delivery both in terms of student performance and measurable gains in efficiency. It is not expected that this will be an easy task.

6. Copyright and IPR

One of the biggest problems which has arisen is that of incorporating copyright material into courseware. The Copyright Licensing Agency has solved this problem with regard to photocopying in education by providing a blanket licence to photocopy within certain guidelines. More work is required, probably at the national level, to negotiate a similar licence for incorporating limited quantities of published material into courseware for educational purposes.

A further complication is the ownership of Intellectual Property Rights (IPR) of developed courseware. One of the benefits of the culture change is the ability of a group of academics to develop courseware modules using common standards. These standards, in turn, allow other academics to modify the material to their own taste. Such an approach goes some way to eliminating the 'not invented here' syndrome; allows obsolete material to be removed and enables new material to be added. It does, however, raise the problem of identifying the true authors (and thus owners) of the material as the courseware evolves. It is thus essential that a methodology is in place to ensure that all authors are properly acknowledged, particularly where the courseware has revenue-raising potential outside UK Higher Education. In this area the expertise and collaboration of publishers will become increasingly important.

7. Summary of Experiences to Date

The University of Nottingham is fortunate in that its senior management is committed to the use of IT throughout the University. This support from above has made the task very much easier. The University already had a CAL element in its overall IT Strategy and the TLTP funding has progressed this strategy considerably. This has meant that the project has met with considerable success at an early stage. The positive experiences of students using CAL courseware have led to some evidence of demand for further development. The main issue which has not yet been properly addressed is the quantification of efficiency gain.

It is felt that institutional projects such as ours have benefited, in comparison with consortium projects, as there has not been the need for inter-institution negotiation of roles, responsibilities and commitment; in addition there has been the advantage of a single well defined IT strategy within which to work.

8. Conclusions

Not only has courseware been developed and delivered in a relatively short timescale, but a general culture change has already begun – staff are more aware of CAL, more inclined to investigate its potential and better able to exploit their ideas. The positive experiences of students using CAL have already led to demand for further development and by the beginning of the 1994/95 session, around 10,000 student hours of CAL developed under the project were being delivered across campus.

It is also becoming clear that CAL enhances the students' understanding of complex concepts across a range of subjects through the ability to simulate a situation and learn at their own pace without worrying about safety or financial implications. Efficiency gains for the university are also significant with less pressure, in particular, on tutorial time and ad hoc consultations.

Nevertheless much work remains to be done. The problems of comprehensive network delivery remain the most formidable, together with copyright issues which require resolution before the goal of flexible and modifiable courseware can become a reality. Nevertheless progress thus far has been most encouraging and, looking to the future, we can already see a need to provide a substantial increase in workstation access to provide CAL in public user areas, which is some measure of our success. Our central aim continues to be the encouragement of both new and existing authors to develop more courseware plus the promotion of the exciting opportunities CAL provides for students across the University.

Reference

DAVIES, P., and BRAILSFORD, T. "Guidelines for Multimedia Courseware Developers in Higher Education", CVCP/USDU, ITTI Publication, May 1994.

Biographies

P.H. Ford

Peter Ford is currently Dean of Science at the University of Nottingham. He is a Chartered Engineer holding a Chair in Information Technology in the Department of Computer Science, and is also Director of the Cripps Computing Centre and the ICL Institute of Information Technology. In addition, he is on the Board of NCC and is a Director of a number of small companies specialising in consultancy and software development. He has previously worked in the aircraft and computer industries as well as in local government. He acts as a consultant on networking issues and information systems development, both in the UK and abroad.

J.E.Graves

Joyce Graves has been Network and Applications Support Manager within the University of Nottingham's Computing Services Department for the last fifteen years. She has over 25 years IT experience covering both data processing and scientific and engineering applications, initially within the Courtaulds Group and later in both end-user and service departments within the University. In addition to her other responsibilities she has overall responsibility for the CAL Support Unit.

The UK Technology Foresight Programme

Marion J Dearlove

ICL Learning, Stevenage, UK

Abstract

This paper summarises the results of the government funded UK Technology Foresight Programme. The programme provided a team of more than 300 UK industrialists and academics with the opportunity to recommend areas for priority action and investment in the UK's 15 major industry sectors. The government decided to undertake the programme after analysing the impact of similar national programmes undertaken by USA, Germany, Australia, the Netherlands and, most successfully, by Japan on wealth creation and innovation. The UK programme was the first to include 'social' sectors – Retail and Distribution, Leisure and Learning and Financial Services and, unlike the other national programmes, focused on both technological and social issues.

The recommendations made by the panels and the Steering Group indicated the need for the innovative use of existing skills and technologies in enabling improvements in the UK competitive position in addition to the development of new ones; the need for close review of existing regulations to ensure that innovation is not stifled; the need for serious improvements in the overall skills of the UK workforce, and for close attention to social issues to improve the quality of life in the UK.

1. Introduction

In the past decade, governments of advanced industrial countries have been faced with increasingly difficult choices in science and technology, as research breakthroughs have expanded the range of new opportunities for advancement, without closing off previous fields of research. This has resulted in an increasing mismatch between the ability of scientists and engineers to develop sound proposals for basic research and the ability of governments to fund them. Governments have, therefore, been forced to look for ways in which to prioritise submissions and to make reasonably informed decisions on funding.

Decisions on priorities require a basis on which they can be set and, as a result of the May 1993 Government White Paper on Science and Technology (17) and a review of the results achieved by the Japanese, Netherlands and German Technology Foresight Programmes (18), the UK Government decided to fund a Technology Foresight Programme in the UK.

Research into the application of 'Foresight' revealed that all previous national programmes had concentrated their attention on identifying improvements in the traditional uses of science and technology, most notably: national security, chemicals, manufacturing, health, transport, materials and energy. The mixed record of attempts made in the 1970s generated considerable scepticism in many countries but the Japanese, who published their first results in 1971, entered into a regular national 5 yearly forecasting cycle which has proved to be extremely beneficial to their economic and social development, so much so that many other countries decided to repeat the exercise or, as in the case of the UK, undertake a Foresight Programme for the first time.

2. The Objectives of UK Foresight

The stated aims of the UK Foresight Programme were:

- to identify those generic technologies which are likely to yield the greatest long term economic and social benefits for the UK in the next 20 years
- to improve communications, interaction and mutual understanding between the scientific community, industry and government departments.

The results of the above would then underpin the key objectives:

- to improve UK competitiveness (20)
- to improve wealth creation potential
- to improve the quality of life of all UK citizens.

3. Sector Selection

Within the programme, the UK Government decided to address the three key areas of UK wealth creation: leisure (12%), financial services (7-8%) and retail & distribution (11%). These sectors had never been considered in previous national Foresight exercises. Their inclusion in the UK exercise has resulted in considerable interest from other countries, specifically in identifying the level of co-operation from within the sectors and in the results generated.

In addition to those listed above, other sectors considered were the traditional areas of: aerospace and defence (3%), transport (8%), energy (5%), materials(3%), manufacturing (22%), chemicals (2%), construction (8%), agriculture (3%); the more technological areas of communication (8%), business processes, information technology & electronics (5-6%); and the more socially oriented areas of health and life sciences (5%), education, natural resources, food & drink (3%) and the environment.

4. Panel Selection

The Programme was initiated in 1993 with the formation of the Steering Group. The role of this body was to provide advice and guidance to the 15 Sector Panels and to analyse the resulting sector reports and establish the underlying generic technologies (16). Their initial task was to select the membership of the various panels. They chose to use a national co-nomination process (19). Known experts were each asked to nominate 10 others. These nominees completed a questionnaire which allowed their levels of expertise in the 15 industry sectors to be assessed. The nominees were then asked to provided 10 nominations.

In 1994, the final list of some 6,000 names was further analysed and those who had been nominated by the highest number of people were then asked to chair the panels and the panels were then selected from the list in consultation with the newly elected chairman and the Steering Group. The remaining nominated candidates became the expert pools allocated to the panels for future consultation.

5. The Foresight Process

In May 1994, each sector panel was targeted with achieving the stated aims and objectives of the programme and set out to:

- establish the sector economic background
- set appropriate targets for growth
- identify priority areas/markets for the UK to improve wealth creation and quality of life
- make key recommendations for action and investment
- specify the initial steps to be taken and by whom.

As part of the background survey, each panel undertook a review of its sector to establish the status quo. This also allowed them to identify what was available in terms of products, services and technologies and to predict likely trends. The areas identified were then assumed to be too short-term to require further investigation and that investment in these areas would come from industry, since the markets already existed. This allowed the panels to focus on requirements for the longer term.

During the investigative period, many of the more 'technological' panels were accused of doing 'what they always do' and approaching the requirements from the technology rather than the perceived need. This was countered by the argument that the technologists were defining what could be done and that the more market oriented panels should be indicating what was required. The IT & Electronics and Communications communities were specifically accused of 'dumping' technologies on the populace without any clear indication of long term side effects. What are the long term effects on society of exposing children to technologies such as multimedia or virtual reality? Will they become introverted or anti-social if distance learning removes the need for the classroom environment? How are computer games technologies affecting society now? The truth is, we do not know. The issue was never fully resolved but the accusation did lead to a more socially directed set of results than was initially indicated by early publications and 'technology in society' research was recommended by a number of panels.

Through a series of meetings, questionnaires, discussions and interviews, the panels were able to formulate a series of ideas which they believed essential to improving wealth creation and quality of life.

On the basis of these ideas, the panels undertook a national survey and a series of regional workshops in which the country's recognised experts in large companies, small/medium enterprises, academia, research establishments and government bodies were asked to give their views on the feasibility of the emerging finding of the panels. The resulting inputs allowed the panels to determine which of the topics were feasible, implementable with or without international co-operation and whether they would be acceptable. These results formed one of the key inputs to the final recommendations.

6. Forming the Recommendations

The final recommendations of the panels covered many wide ranging subjects, in addition to technological advance. During their investigations many of the panels concluded that the real issues were related to society and the need for social change, rather than the development and introduction of new technologies. Consequentially, recommendations were made in many of the reports, and in the report of the Steering Group, that reflected the need for social research and associated action. Other non-technology areas that were the subject of specific recommendations were: new materials; new methods of manufacture, research and development; improved structures at all levels from molecular to major constructions; and environmental and regulatory changes. These areas are considered to be outside the scope of this paper. Those readers who require a comprehensive analysis of all of the issues should refer to the full sector reports (1-15).

The consensus view of the non-technologists was that they had identified the priorities in their sector and that it was now the role of the UK technologists and researchers (like ourselves) to create the products, services and systems required to meet the needs they had identified. It was also concluded by all panels that many of the new markets, business opportunities and 'end' technologies that were required could not be created by one enterprise and would be based on co-operation across traditional company and sector boundaries: IT, Electronics would form closer links with Communications (ITEC), Health with ITEC, Materials with Life Sciences, Biology with Chemistry, Environment with Energy, to name but a few. Co-operation between sectors is seen as being key to the creation of competitive advantage and innovativeness for the UK.

7. Common Themes (16)

Within the sector summaries (see Appendix A) many common themes can be identified. The Steering Group has focused its attention in these areas in constructing its over-arching report and has addressed common themes in three main priority areas:

Creating the environment for investment in the UK by:

- developing the UK workforce knowledge and skill base
- creating the right communications infrastructure
- reviewing policies for long term investment in innovation
- promoting innovation through the regulatory framework
- supporting multi-disciplinary research.

Encouraging UK industry growth by:

- establishing 'centres of excellence' for key industries
- improving the business processes and services through the use of technology
- investing in advanced materials and technologies
- investing in strategic communications and the IT industry
- developing industries based on genetics
- developing multi-disciplined information businesses.

Impacting and shaping society by:

- monitoring demographic change
- analysing the nature of acceptable risk
- monitoring changing patterns of work and leisure
- monitoring the effects of IT on society
- creating a cleaner sustainable environment through technology
- improving our understanding of the human factors in developing markets.

The technologies and skills that underpin these recommendations are many and varied. Many of them exist today in other areas and it is recognised that, the innovative application of existing ITEC products and co-operation across sectorial and organisational boundaries can enable early progress in many of these areas and it is also recognised that new technologies and skills are required and that their early development is vital to the UK's future success.

The information and communications technologies identified as underpinning future industrial and social developments in the UK were:

- Those which allow humans to communicate with machines or with each other through electronic means. Included in this group were: multimedia, virtual reality, information superhighway, speech recognition, image analysis.
- Those which improve design and system integration, most notably: concurrent engineering; rapid prototyping systems; CAD/CAM; requirements and system modelling tools; and sub-system integration technologies.
- Those which improve information management, which include: databases, security, retrieval, etc.; information flows and reservoirs; information security, integrity and protection facilities; and validation tools.
- Those which provide modelling, simulation and prediction of complex systems including molecular design, product design, environmental simulation and behavioural analysis.
- Optical technologies including display technologies, optical sensors, optical information storage, optical computing, optical communication and signal processing.
- Software Engineering with a emphasis on substantial improvements in techniques for creating safety critical real-time systems.
- Telepresence, defined as the design, management, organisation and distribution of information includes technologies such as multimedia, information superhighway to provide services such as tele-medicine, distance learning, entertainment and remote retail.

The finer detail that under lies these definitions can be found in Appendix A in the summary of the recommendations of the 15 panels.

8. Next Steps in Foresight

The government has recognised the need to make progress in the areas identified by the panels and the Steering Group. It has taken on-board the recommendations on regulation and government action and many of the common themes, most notably those related to legislation, education, skills development and health, are already being reflected in ministerial speeches.

It has been agreed that the current Technology Foresight programme will continue through a dissemination phase and a subsequent three year implementation phase.

To ensure action, government funds have been allocated to promote Foresight related activities. £110 million has been provided for implementation which will be managed by the DTI and Office of Science and Technology. Partnership groups, made up from industry and research establishments, will be invited to bid for funding for projects that reflect the recommendations of one or more panels.

9. Conclusions

It is too early to judge the value of the programme to the UK. This judgement will only be possible after the dissemination phase is completed and UK industry and academia have had a chance to evaluate the recommendations.

Initially, the participants were sceptical about the value of the exercise. However, having completed the first phase most are now convinced that the networks resulting from the programme and the priorities identified could be of significant benefit if the implementation phase is properly co-ordinated. Successful implementation could result in UK industry having access to products, systems and services it requires to contribute successfully to the development of the UK economy.

The reports published by the 15 sector panels each contained a series of recommendations. These recommendations were made to government departments and professional bodies but, most importantly, to UK industry and academia itself. In general, the view was taken that government could legislate for wealth and quality of life improvement, and it was encouraged to do this, but only UK industry can create wealth and, by so doing, have the greatest impact on improving the quality of life in the UK.

References

1 Technology Foresight Report 1 – Chemicals. HMSO for the Office of Science and Technology, London, 1995

2 Technology Foresight Report 2 Construction. HMSO for the Office of Science and Technology, London, 1995

3 Technology Foresight Report 3 Financial Services. HMSO for the Office of Science and Technology, London, 1995

4 Technology Foresight Report 4 Health and Life Sciences. HMSO for the Office of Science and Technology, London, 1995

5 Technology Foresight Report 5 Transport. HMSO for the Office of Science and Technology, London, 1995

6 Technology Foresight Report 6 Communications. HMSO for the Office of Science and Technology, London, 1995

7 Technology Foresight Report 7 Food and Drink. HMSO for the Office of Science and Technology, London, 1995

8 Technology Foresight Report 8 IT and Electronics. HMSO for the Office of Science and Technology, London, 1995

9 Technology Foresight Report 9 Manufacturing and Business Processes. HMSO for the Office of Science and Technology, London, 1995

10 Technology Foresight Report 10 Materials. HMSO for the Office of Science and Technology, London, 1995

11 Technology Foresight Report 11 Agriculture, Natural Resources and Environment. HMSO for the Office of Science and Technology, London, 1995

12 Technology Foresight Report 12 , Defence and Aerospace. HMSO for the Office of Science and Technology, London, 1995

13 Technology Foresight Report 13 Energy. HMSO for the Office of Science and Technology, London, 1995

14 Technology Foresight Report 14 Leisure and Learning. HMSO for the Office of Science and Technology, London, 1995

15 Technology Foresight Report 15 Retail and Distribution HMSO for the Office of Science and Technology, London, 1995

16 Report from the Steering Group of the Technology Foresight Programme, HMSO for the Office of Science and Technology, London, 1995

17 Realising our Potential – a Strategy for Science, Engineering and Technology. Office of Science and Technology, Cm2250. HMSO, London, 1993

18 MARTIN, B.R. and IRVINE, J. Research Foresight – Priority Setting in Science. Inter Publishers, London and New York, 1889.

19 GEORGHIOU, I. LOVERIDGE, D. NEDEVA, M. Co-nomination in Foresight. Cabinet Office, London, 1995

20 The World Competitiveness Report 1994. International Institute for Management Development, CERN, 1994

Biography

Marion J Dearlove

Marion joined ICL when STC and ICL merged in 1984. She has a very technical background and has worked mainly in the field of communications systems. Since joining ICL she has operated in the roles of development manager, project manager, bid manager, business development manager and marketing manager. During the period May 1994 to May 1995 she was seconded to the Cabinet Office to act as technical secretary to the IT & Electronics and Retail & Distribution Technology Foresight Panels and was responsible for formulating the final reports of both panels. She has now joined ICL Learning as the programme manager for the Systems College Advanced Programme and aims to apply much of what she learnt during the Foresight Programme to ensure that ICL's senior professional services staff have the skills they need to make ICL successful.

Appendix A UK Sectorial Priorities for ITEC Development

The information below reflects the results of the work of the 15 sector panels and is merely a summary of the relevant sections of the panels published reports. It focuses on the priority areas and technology requirements, reflecting the subset for emerging ITEC technologies only. It must be stressed that this summary by no means reflects the full definition of requirements and that many of the technologies specified in the full set relate to other non-ITEC technology areas.

To improve clarity the following definitions apply throughout:

- a PRIORITY AREA is the area in which research needs to be undertaken or in which a technology can/needs to be applied,
- a KEY TECHNOLOGY AREA is the research or industry from which the required technologies can be derived.

It should also be noted that the summaries are listed in alphabetical order, rather than in publication order with the specific intent of improving location and accessibility.

A1. Agriculture, Natural Resources and the Environment (11)

The priority areas identified for action were:

- integrated ecosystem management for terrestrial, aquatic, coastal and oceanic systems
- alternative energy sources
- animal, microbiological and plant biotechnology and cognate sciences to underpin new products and processes in a wide range of activities from agriculture to pharmaceuticals.

The key technology areas identified were:

- modelling, monitoring, analysis, forecasting and risk analysis systems to underpin environmental research, life cycle evaluation and management, and eco-design.
- robotics in the form of remote sensors and surveillance systems to improve environmental control.
- technologies to support land-fill management, soil remediation, ground water clean-up, and the production and distribution of potable water.

A2. Chemicals (1)

The priority areas identified for action were:

- provision of technically competent manpower at all levels – through the provision of education, training and skill transfer and through an increased awareness of the need for chemicals in UK industry and schools, leading to the development of the skills base for the future.

- a focus on technology diffusion and 'networking' between the industry (major companies and SMEs) and its technology partners and UK researchers.

The key technology areas identified were:

- biochemical technologies
- synthesis, processing and property/structure relationship building in materials and polymers
- robust sensors and measurement systems for the control of processes
- high performance modelling and control systems
- systems to support the control of dangerous substances and environments.

A3. Civil Aerospace (12)

The priority areas identified for action were:

- aircraft design and construction – subsonic, supersonic, commercial transport, rotor craft, turboprop
- alternative propulsion systems
- air traffic management
- airport systems – passenger control, Common User Terminal Equipment (CUTE), environmental protection.

The key technology areas identified were many and varied and included:

- fibre optic and satellite communications, surveillance and navigation systems
- sensors, bio-sensors, surveillance, safety and security systems
- systems for designing factories
- product design systems
- flight management systems
- passenger and freight management systems
- high performance computing for modelling complex systems.

A4. Communications (6)

The priority areas identified for action were:

- addressing the regulatory issues surrounding the provision of the national communications service in the light of increasing convergence in communications.
- addressing the long term requirements for the national communications infrastructure deployment and establishing the relevant technology objectives.

- addressing the need for increased public awareness of the importance of IT and Communications to UK industry, economic performance
- developing the relevant IT and Communications skills in the workforce.
- the Government should aim to become a leading-edge user of IT and Communications and offer public services through electronic media rather than paper.

The key technology areas identified were:

- for infrastructure – digitisation, broad-band networks, mobility and intelligent networks conformant to global standards
- for skills development – access to Internet (or equivalent) from schools with access to interactive educational software, video-on-demand and other educational services
- on regulation for convergence – an electronic services infrastructure will be required that offers end-to-end compatibility of connection, access, authentication, encryption and service provision
- multimedia terminals or relevant subscriber access technologies.

A5. Construction (2)

The priority areas identified for action were:

- the development of a competitive infrastructure with better business processes and more customised solutions from standard options
- better assessment of environmental and social consequences of development
- improvements in education and the promotion of learning networks

The key technology areas identified were:

- information and communications technologies targeted at the construction industry into improve supply chain management and business process design
- virtual Reality for construction simulation and for use within the constructed environment to simulate alternate environments
- modelling systems to allow structural and environmental modelling and testing.

A6. Defence (12)

The report did not specify priority areas and the technologies deemed to be relevant to national security were also omitted. All information deemed to be relevant to national security is available to those who need to know it.

The technologies covered by the report were:

- skills development – IT specialists in systems engineering, systems integration, professional services, contract management, consultancy, analysis, risk assessment, life cycle costs, synthetic environments

- tools to improve optimisation of design, concurrent engineering manufacturing and maintenance
- high performance systems for modelling, simulation and generation of synthetic environments, effectiveness and performance
- sensors, signal processing, data processing and fusion systems
- tools and methods to improve high-integrity, real-time software engineering
- smart structures and skins (for aircraft and missiles) – with embedded sensors, electronics and actuators
- high performance guidance and control systems
- command, control, communications and intelligence systems
- robotics and automation systems
- non-lethal weaponry and electronic warfare systems.

A7. Energy (13)

The priority areas identified for action were:

- supply – new energy sources, emission and cost reduction, exploration, decommissioning
- transport – improved methods of transport and storage
- conversion – improved methods of converting primary sources to delivered fuel
- Usage – efficiency, conservation, customer management, payment systems, usage monitoring and control and improved energy services.

The key technology areas identified were:

- advanced information technology, software, process control systems and high performance communications systems
- high performance modelling systems
- image analysis
- sensors and surveillance systems.

A8. Financial Services (3)

The priority areas identified for action were:

- the definition of a regulatory framework for an appropriate telecommunications network to support the development of and access to secure electronic financial services
- improvements in the UK skills base in the areas of numeracy, literacy and IT literacy at all levels through improved provision of IT, network access and appropriate training facilities

- improvements in security, surveillance, fraud detection and prevention.

The key technology areas identified were:

- a communications infrastructure that supports the provision of secure financial services that includes multimedia access, service and subscriber authentication, encoding, encryption and transaction monitoring
- fraud detection and prevention services based on biometrics, artificial intelligence techniques, neural networks for the analysis of behaviour patterns, sensors
- secure technologies for smart cards
- customer information management systems
- on-line financial service packages with client usable interfaces
- development of new educational packages and services to improve skills levels in schools and the adult population.

A9. Food and Drink (7)

The priority areas identified for action were:

- improvements in the safety of microbiological, chemical and physical food and drink processing and distribution
- exploitation of the growing capability of biotechnology to modify the properties of agricultural products
- improvements in the understanding of food production through multivariate modelling
- increasing multi-disciplinary research into food production, establishing the relationship between health and diet and the understanding of psychology of consumer choice.

The key technology areas identified were:

- high performance sensors to detect changes in the materials during food production
- high performance systems for genetics research, molecular modelling and biotechnology developments
- databases, information networks and control systems for food production.

A10. Health and Life Sciences (4)

The priority areas identified for action were:

- integrative biology – research programmes which integrate molecular biology and genetics with cell and tissue biology, and whole organism studies

- neuroscience and the cognitive sciences – research into progressive degenerative disease and non-specific, age-related decline
- applying research into disease at the genetic, molecular and cellular levels to develop new generations of diagnostics
- ageing and disabling degenerative disease
- drug creation, testing and delivery.

The key technology areas identified were:

- advanced recombinant technologies includes research into metabolic pathways, metabolic engineering and applications in the biological manufacture of industrial products
- high performance systems (including computers) for genetics research, modelling, data collection and analysis
- infrastructure for exploitation and development of life sciences and clinical medicine
- communications infrastructure that allows the access and transfer of clinical information, in real time, between interested parties
- technologies for sustaining a reasonable quality of life for the elderly infirm.

A11. IT and Electronics (8)

The priority areas identified for action were:

- promoting investment attractors from UK and overseas by ensuring that the UK has a highly educated workforce, first class research and access to leading edge technologies
- developing a coherent Information Superhighway policy, building on the strengths of the key players, to provide a leading position for the UK in the implementation of a national Information Superhighway
- encouraging the development of new multi-disciplinary, content-based electronic businesses including tele-medicine, remote learning, environmental monitoring and control, financial services, 'edutainment', leisure products, home shopping, etc.
- IT embedded in cars, televisions, washing machines, etc. is becoming increasingly common. The skills required to develop embedded systems are very different from those currently taught in UK universities to Computing Science students. The UK must improve its ability to develop the skills required to create embedded IT systems
- encouraging co-operative developments between the IT sector and other industry sectors to address specific requirements
- networking facilities for all educational institutions (at all levels) and the development of educational packages for both classroom access and distance learning.

The key technology areas identified were:

- a globally compatible communications infrastructure that offers seamless end-to-end connection
- a communications infrastructure that also provides secure access between subscribers and services
- standards for privacy, security, access and IPR protection in the Information Superhighway environment
- microtechnology – next generation semiconductors, micro-machines, etc.
- display, optical and acoustic technologies – providing the platform for multimedia developments
- virtual reality – improved visual and sensory systems providing the platform for modelling, simulation, education and entertainment
- distributed information management – providing control and navigation for networked systems such as those offered on the Information Superhighway
- co-operative developments between media specialists, information technologists and communications experts to develop multimedia systems and packages for the home, the work place and the educational environment
- sensors, bio-sensors and surveillance systems
- speech and Image processing
- security and authentication
- formal methods and modelling
- high capacity/performance computers
- radio and Digital Signal processing systems.

A12. Leisure, Learning and Media (14)

The priority areas identified for action were:

- improvements in the skills of the UK workforce
- developments in distance learning and learning systems at all levels
- convergence in arts and sciences (as witnessed in the recent developments in the animation and fashion industries) resulting from technology usage
- UK could become a major provider of learning materials on a global basis
- improved application of technology to sports equipment
- networked sports fixture and partnering services
- heritage exploitation

- continuous improvements in the music, film, special effects, animation, games and television industries.

The key technology areas identified were:

- technologies to deliver and receive education and entertainment in the home, work place and training centre
- multimedia education packages developed for delivery on CD-ROM or across a network
- on-line libraries, databases and educational services
- expert systems to assist and support self-learning and practice in education and sport
- multimedia leisure and education packages (for CD-ROM and network distribution) based on UK heritage, galleries and museum collections, and tourist sites
- virtual reality simulations
- continuous improvements in the UK leisure software industry (games, music, etc.)
- technologies to support the increasing requirements of special effects, animation, film, etc.

A13. Manufacturing and Business Processes (9)

The priority areas identified for action were:

- increasing the emphasis on business processes – research into best practices, together with diffusion of practices throughout industry
- national infrastructure for improved competitiveness – by taking action in the areas of education, transportation, deregulation and finance
- improving the supply chain – networking between customers, partners, suppliers and manufacturers
- increasing technology transfer and investment throughout industry.

The key technology areas identified were:

- processes, plant and equipment which meet future environmental needs
- advanced sensors and controls
- modelling, simulation and visualisation for technical and business use
- integrated IT systems supporting effective business processes
- improved process plant productivity, quality, repeatability and effectiveness
- improved processes to use new materials effectively

- educational facilities and packages to promote improved IT skills in the UK workforce.

A14. Materials (10)

The priority areas identified were:

- the continuous improvement of existing materials and processes
- materials and processes which improve the environment or have specific applications, e.g. in health care.

The key technology areas identified were:

- modelling systems that can be used to reflect material composition, structure and process performance to end-product performance
- improvements in sensor materials and devices and automated process control
- processing technologies which improve the environment
- weight-saving technologies for specific applications
- processing techniques for high temperature super-conducting materials
- development of materials for IT, Communications, durable joining techniques and surface engineering.

A15. Retail and Distribution (15)

The priority areas identified for action were:

- promoting the development of globally compatible retail technologies by addressing such issues as inter-operability and use of open standards
- improving the end quality of products by improving the ability to monitor their progress through the supply chain and the conditions in which they are kept
- improving the management and security of customer information
- improving access to global markets
- establishing and improving remote retailing facilities through the Information Superhighway
- improving the face-to-face retailing experience
- gaining a better understanding of the appropriate use of technology in retailing
- improving the skills of the UK workforce through education, adult learning and skills transfer
- establishing coherent policies for land use, transport
- gaining a better understanding of demography.

The key technology areas identified were:

- a national Information Superhighway that offers end-to-end compatibility and secure global access. The system should include encryption, authentication and access technologies, IPR protection and payment systems, Information management services and navigation services
- smart tags
- smart cards – for customer and financial information, and system access
- secure customer information management systems
- improved EPoS systems which offer the opportunity for increased product and customer management
- multimedia hardware and software
- home shopping terminals that provide the general public with an easy-to-use interface to retailing services
- improved distribution technology for product and vehicle tracking and product and load distribution.

A16. Transport (5)

The key priorities were to establish the following showcase projects:

- the Informed Traveller – designed to provide integrated real-time information, ticketing, booking and payment seamlessly across all passenger transport modes
- the Foresight Vehicle – designed to stimulate the UK automotive supplier base to produce vehicles which satisfy increasingly stringent environmental requirements while meeting mass market expectations for safety, performance, cost and desirability
- Clear Zones – designed to provide a focus for the many disparate activities related to creating more liveable urban centres.

The key technology areas identified were:

- pattern processing and recognition technologies
- safety critical systems
- high performance multimedia systems that provide access to high performance information storage and retrieval systems
- Information Superhighway infra-structure and super-structure
- accurate location systems
- fuel efficient, low emission power plants and energy recovery systems.

Making the Internet Safe for Business

P. V. McMahon

ICL Enterprises, Reading,, Berkshire, UK

Abstract

With the intense media interest in the Internet, many concerns are being expressed about connecting an organisation's computing infrastructure to what is essentially an unregulated and insecure world-wide network. This paper identifies the threats to business information and IT systems posed by the Internet, discusses the types of solution which can be implemented to address these concerns, and shows how ICL's Access*Manager*™ provides a strategic foundation for an enterprise's secure exploitation of the Internet's facilities.

1. Introduction

This paper describes the security problems in the current deployed version of the Internet Protocol Suite (IPS) then discusses how these threats can be countered by different protection mechanisms such as protected enclaves and secure protocols. A brief overview of the advantages and disadvantages of each class of protection mechanism is given, and current industry trends are identified.

It then discusses how Access*Manager*™ [ICL, 1995] supports implementation of a secure network perimeter, as well as meeting internal enterprise access control requirements; and how planned Access*Manager* firewall support can be exploited.

2. Security Limitations of the Internet

2.1 Problems

The IPS clearly reflects its origins as a system for linking a cooperative group of researchers by its lack of intrinsic defences against malicious attacks. Most existing IPS implementations use static passwords for authentication of users, and addresses for authentication of systems.

Hence both IPS infrastructure (routing and name servers), and end-to-end communications (application protocols) are vulnerable to such threats as eavesdropping, masquerade and undetected modification of packets.

Furthermore, as the Internet user has traditionally not been charged on a per-usage basis, there are no reliable mechanisms for ensuring accountability of user actions.

The Internet also directly contributes to its own subversion as it provides a communications channel through Usenet[1] and mailing lists for knowledge of vulnerabilities to be widely disseminated.

2.2 Commercial Implications

Both the Gartner and META groups have commented that the current justified concerns about Internet security have a chilling effect on the otherwise desirable migration towards widespread use of the Internet for business transactions.

The existing use of commercial value-added-networks (VANs) for inter-enterprise EDI and messaging relies on the network service provider for security guarantees of business information not offered by Internet service providers.

The large commercial involvement in the many current initiatives on Internet security (e.g. Internet Engineering Task Force (IETF), CommerceNet, OSF Research Institute, CEC ACTS etc.) is motivated by a desire to tap into the large and growing market for electronic commerce which is currently dominated (and limited) by VANs.

The need for an inter-enterprise network which is safe for business will drive the development of a secure Internet.

3. Approaches to Securing the Internet

3.1 Protected Enclaves

Given the limitations of security in the current IPS, the most popular approach to achieving secure Internet access currently is for organisations to drastically curtail Internet connectivity for their users, and therefore limit the exposure of their business information to hacking.

This requires organisations to set up internal networks as "protected enclaves" [Shirey, 1994] which are surrounded by secure perimeters, or "firewalls" (see section 4).

The advantages of this approach are that existing insecure IPS systems can be used internally to access specific external IPS applications. Secure incoming access is also possible, but this requires additional strong authentication checks such as those provided by the secured transport relay function in version 5 of the "SOCKS" protocol [Leech, 1995], [McMahon, 1994].

---

[1] Usenet is the name of a set of independent, originally mainly academic computers that exchange news material in the form of newsgroups labelled with their topics. There is no central authority; Usenet traffic may be but does not have to be carried by the Internet.

The disadvantages of this approach are potential limits on which applications can be supported; and lack of protection within the protected enclave. It is also necessary to enforce internal controls effectively to prevent independent unauthorised firewall-bypassing access to the Internet from subverting the security policy, and compromising the protected enclave.

Internet Firewalls are further discussed in section 4 of this paper.

3.2 Secure Protocols

The core Internet security problem is that existing IPS authentication and data protection is highly vulnerable to attack and makes masquerade attacks very easy.

The need to address this problem has resulted in efforts by the Internet Architecture Board to define an approach for a new security infrastructure [Braden et al, 1994]. Practical work is already underway to make the security of the deployed IPS, and IPS applications, resistant to attacks. This is happening both at the application and network layer.

End-to-end protocols such as **snmp**, **ftp**, **telnet**, **imap**, and **ppp** are being developed to include cryptographically strong security services – minimally authentication, but also integrity and, optionally, confidentiality. Work is also commencing in the IETF on session layer security and WWW transaction security based on proven commercially available technology [Netscape, 1995] and [Rescoria & Schiffman, 1994].

In addition, in order to facilitate a consistent approach to end-to-end security in applications, a common authentication technology is being developed in the IETF which is currently based on Kerberos V5 [Kohl & Neuman, 1993], and GSS-API [Wray, 1993]. Related work in Europe addressing some of the limitations of Kerberos V5 has been implemented in the SESAME project [Parker & Pinkas, 1995] and [McMahon, 1995].

At the same time, a standard for secure IP is being defined. When deployed this will remove the need for most of the point solutions which are currently necessary in applications. A set of generic security format specifications [Atkinson, 1995a], [Atkinson, 1995b], and [Atkinson, 1995c] have been defined, together with a required basic set of cryptographic protection mechanisms [Metzger & Simpson, 1995] and [Karn et al, 1995]. However, widespread deployment of TCP/IP implementations conformant to these specifications is many years away - most likely as part of IPv6, the next version of the Internet Protocol Suite.

Store-and-forward protocols must, however, include security with the data object – and the PEM/PGP systems are being increasingly used, and are available commercially.

3.3 Internal Controls

Numerous surveys and industry analysts' reports have confirmed that most security breaches are perpetrated by insiders rather than external hackers.

Hence, enterprises must, of course, not only employ appropriate security measures against a hostile Internet, but also have effective internal access controls and administrative procedures.

3.4 Manageability

For the enterprise's IT security to be coherent and efficient to manage, Internet security measures should be part of a single administrative view, rather than disjointed and uncoordinated.

However, many "protected enclave" solutions today fail to meet this goal – as internal access controls are not provided.

Similarly, those Internet application security protocols which are currently defined do not support a consistent management view. Hence while the basic technology for security is being engineered within the IETF, there is still a significant amount of coordinating infrastructure needed in order to harness this technology, and make its costs of ownership commercially acceptable.

4. Firewalls

4.1 Definition

A firewall is a means of establishing and defending a protected enclave (as discussed in the previous section).

"Firewall" is typically used as a catch-all term for a set of hardware and software security facilities which protect the business information and computer systems within an organisation's network from unauthorised access and external attacks.

More precisely, a firewall is generically defined in the emerging Internet Security Architecture [Shirey, 1994] as a communications relay which restricts traffic (according to security policy). This filtered relay function may be at any layer of the communications stack.

While traditionally used to apply to systems which counter external threats, firewall infrastructure can also be used to prevent unauthorised use of external computer systems from within an organisation.

4.2 How can Firewalls be Guaranteed to catch all Incoming Attacks?

For firewalls to be effective, physical isolation of the internal network perimeter must be achieved by forcing all incoming and outgoing traffic to traverse a single, controlled point of entry.

The firewall may be a combination of:

- one or more gateway servers (sometimes called bastion hosts)
- one or more routers.

The firewall polices both incoming and outgoing connections. This is because even though the primary source of threat is external, outgoing traffic results in applications being started which permit incoming traffic – hence outgoing traffic must be controlled to guard against external threats.

4.3 Firewall Components

Using the [Cheswick & Bellovin, 1994] model, a firewall may comprise any or all of:

- application-specific gateway for mediating outgoing and incoming traffic (e.g. mail gateway, http proxy running on a gateway host)
- packet filter - a hardware or software device which permits only those IP packets matching specific criteria to traverse a firewall (e.g. router)
- a TCP/IP relay, or circuit gateway (e.g. SOCKS version 4 of which relays TCP and is being extended in version 5 to relay UDP)

4.4 Firewall Filtering Policy

A firewall may use some or all of the address, port, and authentication information associated with the packet or application-level protocol.

Different considerations are associated with incoming and outgoing traffic:

- Incoming

  Any attempt to establish a connection from an external source should be prevented unless policy explicitly permits it.

  External packets for existing connections can be permitted, as if they are not associated with an existing connection, they will be rejected by the internal destination host.

  For some applications, such as mail, authentication of the external packets is not required.

  Standard **ftp**, **rlogin**, and **telnet login** protocols are usually password-based and are highly vulnerable to hacking.

  In general, authentication mechanisms which are not vulnerable to attacks are required to permit such services as to be used by those outside the firewall to call in.

- Outgoing

  The concerns here are with the amount of time and money which employees can spend accessing external services and "browsing" the Internet.

  For application gateways, a usable and appropriately secure means of insiders accessing the proxy servers is required in order to enable charging and accountability.

4.5 A Continuing Role for Firewalls

The strong industry support for IPS, and the prospect of a paradigm shift in usage as more and more devices become addressable, all lead to some pressures for deployment of IPv6.

Implementation of security is mandatory in IPv6, and will, together with an Internet-wide secure DNS and KMP, enable any two IP hosts to communicate securely.

The trends towards use of third party Internet Service Providers, innovation in application development, and increased user requirements for connectivity will all work together to push security back into the host, and will eventually remove need for firewalls purely for external security reasons.

Hence, the proliferating firewalls, gateways, and proxy applications of today are likely to be transitional mechanisms to cope with the current inherent security limitations of IPS, but will not be security solutions forever.

However, until every IP node is secure, there is a critical need for firewall solutions to provide confidence to those who need to use the Internet. Furthermore, in addition to their current role in intrusion protection, firewalls enable organisational access control of Internet use from inside, and enable benefits of centralised security administration to be realised.

Therefore firewalls can be expected to have a continuing use within large enterprises, but with their role shifting from being the sole line of defence against a hostile Internet, towards being a supporting component within a wider Enterprise network security infrastructure.

5. Access*Manager* and Internet Security

This section identifies the facilities provided by Access*Manager* which support secure use of the Internet, and shows how the product can be used now to support the establishment of a protected enclave, and how planned enhancements to Access*Manager* will extend the solution capability of Access*Manager* to reduce management overheads, and improve security.

5.1 Functions

The role for Access*Manager* in the Internet environment is:

- Single sign-on

  To provide SSO for internal users accessing application proxies in order to connect to the outside world. This enables any complex scripting or authentication dialogues to be performed on behalf of the user.

- Internal Access Control

  To permit accountable access control, so that only authorised users can access external services

- Firewall authentication server for remote user sessions

  To provide a secure user-host authentication between a user outside the network perimeter, and a User Sponsor running **behind** the network perimeter

- Secure relayed TCP/IP

  To support secure access from outside a firewall to "legacy" unchanged TCP/IP applications within a firewall via a secure transport relay.

- Secure access for applications

  To support secure application protocols to give end-to-end security from clients to servers.

5.2 Protected Enclave Internet Security using Access*Manager*

Access*Manager* can be used now in conjunction with a filtering router to provide manageable security for existing IPS systems.

A bastion host can be set up as part of a firewall which will reject standard UNIX logins, but will permit logins via an Access*Manager* User Sponsor using an appropriate[2] authentication mechanism which can then provide access to the internal network to authorised users. This use of Access*Manager* as an authentication server for filtering incoming user access is scaleable as users always access their home Person Server to authenticate – and is therefore suitable for firewalls in large organisations.

Mobile or other external users can therefore use Access*Manager* to securely (i.e more securely than UNIX login) login over the Internet, or dialup, to their home services.

Access*Manager* is also being extended to include a secure transport relay which can be placed on a bastion host to provide per-packet authentication in order to give an enhanced level of security.

The use of Access*Manager* at the gateway, and within the protected enclave provides a complete integrated security solution delivering both secure external access for remote users, and single sign-on and access controls for internal users accessing external services.

5.3 Application Security using Access*Manager*

Access*Manager* currently supports a profile of the industry standard GSS-API which provides security services necessary for securing applications to enable client/server authentication, and to ensure the integrity and privacy of their data.

The emerging secure Internet protocols will require secure association services to protect applications or packets. This will require both secure association management facilities and their integration into applications and clients, or communications gateways.

Support for SESAME/Kerberos distributed security technology in the next major version of Access*Manager* will provide the necessary infrastructure for end-to-end security in application protocols. As well as securing applications for internal users, it will permit applications to be used securely by authorised external users. Furthermore, unlike the traditional password-based (or address-based) security, secure association management does not require

---

[2] "appropriate" means that AccessManager will be configured to permit only logins under roles which use authentication mechanisms which provide protection against replay and masquerade attacks.

maintenance of user passwords and accounts at every system being accessed, and therefore cuts access control management costs, and risks of system compromise.

5.4 The Internet and Access*Manager*

Access*Manager* can, therefore, be used to achieve two distinct strategic objectives:

- Manageable access control within a protected enclave

  Access*Manager* secure association management can deliver controlled and accountable access from users within an organisation to authorised external systems without the need to set up and maintain lists of users or their machines at a gateway system

- Direct secure use of the Internet

  Access*Manager* secure association management can support end-to-end security of IPS applications so that only authorised users can access enterprise data.

Access*Manager* secure association management can, therefore, make a protected enclave more manageable and, through support of Internet security protocols, eliminate the need for a protected enclave gateway without compromising security.

6. Conclusion

The security of the Internet is one of the most visible issues in today's IT environment.

With the significant vendor and business interest in seeing the correct outcome, market pressures will force the incompatible and unsatisfactory solutions of today to converge within the next few years towards a common network security authentication technology, and a single messaging security system.

Access*Manager* is unique in providing an integrated framework for secure use of the Internet by enterprises now as a natural extension of the access controls supported for internal systems; and in giving a forwards path for supporting the secure Internet of the future.

References

ATKINSON, R.(a) "Security Architecture for the Internet Protocol", RFC 1825 (Proposed Standard), Internet Engineering Task Force, August 1995

ATKINSON, R.(b) "IP Authentication Header", RFC 1826 (Proposed Standard), Internet Engineering Task Force, August 1995

ATKINSON, R.(c) "IP Encapsulating Security Payload (ESP)", RFC 1827 (Proposed Standard), Internet Engineering Task Force, August 1995

BRADEN, R., CLARK, D., CROCKER, S. and HUITEMA, C. "Report of IAB Workshop on Security in the Internet Architecture, February 8-10, 1994", RFC 1636 (Informational), Internet Engineering Task Force, August 1995

CHESWICK, W.R. and BELLOVIN, S.M. "Firewalls and Internet Security", Addison-Wesley, June 1994

KARN, P., METZGER, P. and SIMPSON, W. "The ESP DEC-CBC Transform", RFC 1829 (Proposed Standard), Internet Engineering Task Force, August 1995

KOHL, J.T. and NEUMAN, C. "Kerberos Network Authentication Service, RFC 1510 (Proposed Standard)", Internet Engineering Task Force, September 1993

ICL "Product Description ACCESS MANAGER 300 for Sun", Product Document PD-AM-0005, ICL, March 1995

LEECH, "SOCKS V5, draft-ietf-aft-socks-protocol-v5 (Internet Draft)", Internet Engineering Task Force, July 1995

MCMAHON, P.V. "SOCKS V5 GSS-API Authentication Mechanism", draft-ietf-aft-gssapi (Internet Draft), Internet Engineering Task Force, November, 1994

MCMAHON, P.V. "SESAME V2: Public Key and Authorisation Extensions to Kerberos", Proceedings of ISOC Symposium on Network and Distributed Systems Security, IEEE, February 1995

METZGER, P. and SIMPSON, W. "IP Authentication using Keyed MD5", RFC 1828 (Proposed Standard), Internet Engineering Task Force, August 1995

NETSCAPE, "Secure Sockets Layer Protocol – Version 3", Netscape Communications Corp, June 1995 (also draft-hickman-netscape Internet Draft)

PARKER, T.A. and PINKAS, D. "SESAME V3 – An Overview", SESAME Project, May 1995

RESCORIA, E., and SCHIFFMAN, A. "The Secure Hypertext Transfer Protocol", Enterprise Integration Technologies, December 1994

SHIREY, R. "Internet Security Architecture, draft-irtf-psrg-secarch" (Internet Draft), Internet Engineering Task Force, November 1994

WRAY, J. "Generic Security Service API – C Bindings", RFC 1509 (Proposed Standard), Internet Engineering Task Force, September 1993

Biography

Piers McMahon graduated from Trinity College Dublin with an honours degree in Computer Engineering. He joined ICL in 1988, and has worked in software development with a specialisation in distributed security. He has contributed to a number of CEC-funded research projects including SESAME and SIPE, and the Technical Demonstrator Programme of the UK MoD. He is currently Technical Strategist for the Access*Manager* product being developed by ICL and EDS, and chairs the Security Working Group of the X/Open consortium.

Developing Financial Services Kiosks

Mike Shaw
ICL Enterprise Technology, Bracknell, UK

Abstract

At the moment, the only way for a prospective customer to get full information about, or to buy, virtually any financial product, is to interact with a member of staff from the financial services organisation. Such customers do not necessarily want full automation but rather the opportunity to move through the buying process on more of a self-service basis. Multimedia 'kiosk' systems provide an opportunity to automate many aspects of the buying process whilst providing higher levels of customer service and satisfaction.

Kiosks are based on personal computing technology; however, as compared with personal computing, the target users of kiosks and their expected skills are fundamentally different. Personal computer applications and infrastructure software tend to need changes or enhancement for self-service kiosk use.

There are many reasons for kiosk projects failing to meet their aspirations fully. These often result from unclear objectives or lack of defined critical success factors for the project at the outset. This paper describes a process for kiosk systems development that contributes towards ensuring project success and minimisation of risks. The process follows OPEN*framework* methods and includes business considerations, physical and environmental design aspects, and the architecture.

1. Introduction

Multimedia 'kiosks' are increasingly populating the 'High Street', shopping malls and airports as many organisations try to derive competitive advantage and provide higher levels of service through use of technology. A kiosk is typically a form of multimedia personal computer mounted in a specialised housing and incorporating touch screen technology. The financial services sector was one of the first to automate the customer interface with the introduction of Automatic Teller Machines (ATMs) in the 1970s; multimedia kiosks extend the self-service approach by providing information on products and services leading to a selling opportunity.

The self-service nature of a kiosk offers a new, "no pressure", approach to assist the sale of financial services. By presenting a comprehensive picture, the kiosk can transform the preliminary rounds of fact-finding associated with the more complex financial products from a chore into an easy and enjoyable process. Staff can now assume a new role as advisors and facilitators in a purchasing decision.

Financial kiosks are normally located at the branch; however; by the inclusion of video conferencing, the service can be extended to non-branch locations.

This then is the background to the formal processes and methods developed for financial services kiosk projects that are described below.

Figure 1 shows an example of a financial services kiosk developed in conjunction with the UK-based Nationwide Building Society.

Figure 1 Nationwide Building Society Kiosk.

The potential benefits offered by this new generation of customer self-service terminals or 'kiosks' include:

- releasing branch staff from mainly routine tasks of information provision, leaving them more time to provide the high-value element of the service
- increasing significantly the marketing impact when presenting the customer with a choice of products and services through the use of interactive multimedia
- projecting a high profile, progressive image as a leading retailer of financial services
- rapid introduction of new products without long and costly branch training programmes

- extended hours of availability through 24 hour lobby access and video-conferencing
- positioned to exploit the predicted multimedia home 'shopping' market.

In summary, a kiosk offers: a method for controlling costs, enhancing the role and effectiveness of branch staff, and, most importantly, improving the quality of service to customers.

2. Developing Financial Services Kiosks: the Process

2.1 Overview

The key to profitable development is maximising investment while controlling cost in an environment of change, where there is a need to provide an ever increasing range of products and services while maintaining an ever improving level of perceived customer service. Proven techniques and processes within an overall structured approach help to ensure these objectives are achieved. The kiosk development process uses and is a microcosm of OPEN*framework* Methods for change described in [Hutt, 1993a].

2.2 Process Steps

There are many facets of a kiosk delivering multimedia-based customer services, all of which need consideration at the inception of a project for it to be successful; even then there are surprises in how customers react to specific aspects. Customer reactions call for an iterative approach with continuous behavioural, ergonomic and usability testing at stages throughout the project.

OPEN*framework* Methods for change (see figure 2) identifies four distinct stages in the kiosk development process: *Direction setting* concentrates on the business and strategic aspects of the project, *Design* formulates the technical solution and approach to meet the business objectives, *Implementation* includes the development, build and system testing phases whereas *Deployment* includes pilot and volume installation phases. Although the techniques described here follow a logical sequence, the process is continuous with each stage feeding back into earlier stages. The *Marketing* programme runs in parallel and has defined relationships with each stage that are described later.

Within this framework there are four distinct threads in the kiosk process; the physical environment including the design or procurement of the kiosk enclosure itself, the technology architecture, the application architecture and the marketing programme.

2.3 Direction Setting

Direction setting draws on relevant strategic information from the enterprise setting the project in the context of overall enterprise goals. It divides into the four elements highlighted in figure 2. *Capturing the vision, Quantifying business benefits* and *Evolving business strategy* are handled through the kiosk business planning process whereas *Evolving technical strategy* is handled by a process in its own right.

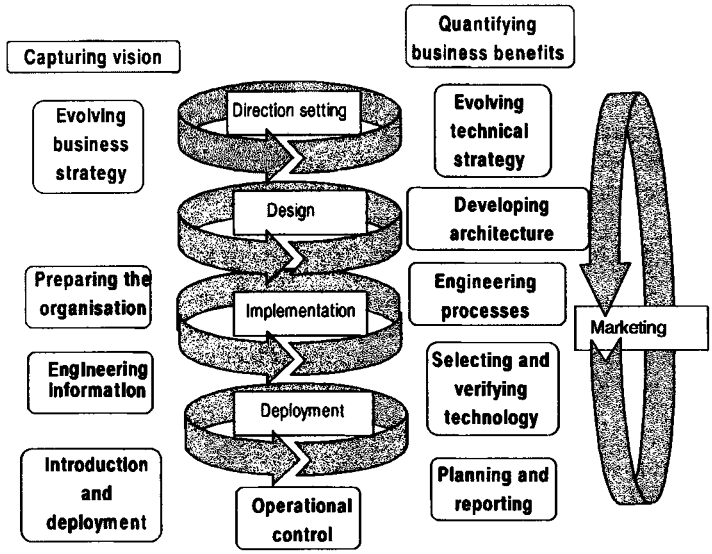


Figure 2 OPEN*framework* Methods for change with kiosk enhancements

Capturing the vision is concerned with the strategic directions of the enterprise and establishing the kiosk project within those overall strategic goals, for example, impact on branch re-design or use in alternative sales channels. The vision allows potential areas of reusable development to be identified in later stages and is a key input to prioritisation.

Quantifying business benefits is probably the most important aspect of this stage. Improved profitability and customer service are two key factors to be evaluated. Time to market for new products can be reduced by exploitation of multimedia kiosks. There may be other business-related research benefits to the enterprise which qualify the pilot phase of implementation.

Evolving business strategy develops the requirements, objectives and critical success factors for the project and, in particular, focuses on those issues that will be relevant in later stages of the process. For example, these include, the target users, 'sit down' or 'stand up' kiosk formats and hours of operation. Financial product suitability and priorities for implementation on the kiosk are determined or confirmed at this stage against marketing input data, for example, the size of market, demographics and competitive positioning.

Evolving technical strategy includes technological requirements, the kiosk housing unit and the overall environment in which it will be used. Levels of integration with existing systems are determined for further consideration during *Developing architecture*. The impact of new and emerging technology within the life cycle of the project is also assessed, for example, intercepts of price/performance improvement curves governing cost-effective timing of volume deployment. Considerations also include core competencies to be developed internally or out-sourced by the enterprise in support of a kiosk programme. This governs the creation of the project organisation for subsequent stages. At this stage, additional project partners and sub-contractors are identified.

2.3.1 Business planning process

Business planning is performed in two steps. The first step is a preparatory document; the business workshop brief, a study report capturing information on the three business related elements of direction setting. This is followed by a decision conference; the business planning workshop to resolve issues raised in the brief, to ratify aspects of the document and to agree project direction.

The purpose of the business workshop brief is to produce an objective assessment of the proposed programme as input to the planning team prior to the business planning workshop. The study is undertaken by a small team, first identifying sources of information and then collating and analysing the data in order to compile the brief.

The planning team typically comprises managers representing: marketing, financial products, branch design, premises/environment, branch/channel management, IT strategy and kiosk provision. The marketing role is to provide, for example, demographics, customer profiles, and target markets; the financial product manager's input covers how products or services can be positioned within the kiosk context and how these or new products are expected to develop within the planning horizon. The kiosk must be designed to fit within the branch architecture and the involvement of channel management at this stage is vital to successful deployment. IT strategy is represented so as to consider impact on and integration with existing systems. IT strategy also covers requirements for new development which are taken as input to the following technology workshop step. Lastly, a workshop facilitator is required.

The kiosk programme then needs to be represented and driven by a microcosm of the organisation. Ideally, the planning team members would ultimately form the central core of the programme's steering committee.

The business planning workshop objectives are:

- to provide a common level of understanding across the enterprise and project partners
- to agree strategic objectives and positioning in market place over a period of typically 3 to 5 years
- to identify critical success factors:
    - impact on customer service
    - impact on profits/costs
- to scope the programme in terms of business needs and priorities, thus providing the basis for detailed project planning
- to define key business applications or services to be developed covering:
    - market differentiation

- functional requirements
- levels of integration
- priorities

Output from the business planning workshop is a clear statement of business and programme objectives, business requirements for the applications and a clear indication of priorities allowing the business strategy for the programme to be documented.

2.3.2 Evolving technical strategy

The technical strategy is the response to fulfilling the business imperatives and addresses technical requirements. For kiosk systems, this divides into three components: environmental, technology and application. At this stage the top level systems design is produced showing how the product might look now and how it could evolve within the planning horizon. The process of *Evolving technical strategy* is shown in figure 3, of which the first iteration is achieved through a technology workshop.

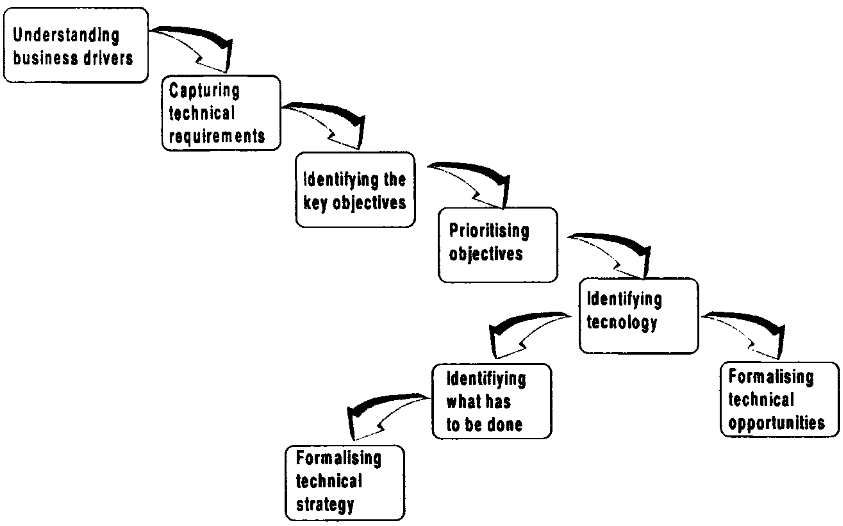


Figure 3 Evolving technical strategy

2.3.3 The technology workshop

The technology workshop is designed to capture and prioritise the technical requirements leading towards the developing architecture stage. Considerations for this workshop are: peripheral specification, personal computer and operating system selection, size of display unit, networking requirements, systems management requirements, top level systems design and information flows, inventory of existing or planned developments that could be reused in this development, integration with any existing systems and input on preferred products, standards and tools.

The technology workshop is attended by representatives of IT strategy and development, kiosk design or procurement and business marketing. The output is a formal statement of the kiosk technical strategy which is compiled and published subsequent to the technology workshop and which becomes a reference document for the design teams.

Peripheral specification is a vital input, any kiosk housing design and device selection is an early activity in the next stage of design.

The design workshop is facilitated using appropriate techniques, for example, House of Quality. See [Combridge, 1992]

2.4 Design

At this stage the design moves forward on each of the three components: environment, technical architecture and application architecture. An architecture is a framework of functional components, standards, and processes in which there is human involvement. When integrated together, the architectural components form the kiosk solution. Developing architecture is described in [Hutt, 1993c]. Design and implementation of the physical kiosk and its environment are described in section 3. Project planning and resourcing of the activities are further parallel actions. Project costs and timescales are constantly reviewed. Completion of the design stage is a major check point for costs, timescales and priorities.

2.4.1 Developing technical architecture

On the basis of the output from the technology workshop, the next activity is to specify elements of the technical architecture. A reference model for elements of the technical and application architecture is contained in [Brunt & Hutt, 1992]. The key elements are:

- hardware and software platforms, peripherals and infrastructure software
- local and wide area networking interfaces, protocols and gateways
- information management databases
- systems management environment

2.4.2 Developing application architecture

In parallel with technical architecture development, the application architecture is produced, focusing on:

- application structure and modules
- application functionality to meet business objectives
- applications development environments and tools
- user interface standards

2.4.3 Analysis of customer behaviour

For the technology to perform, it must be accurately targeted: crucial to the success of the project is an accurate understanding of the behaviour of financial services customers and branch staff at each stage of the buying process. The buying process stages include:

- browsing
  - seeking information
  - reviewing alternatives/options
  - learning
  - increasing confidence in making a decision
- evaluating
  - evaluating alternatives
  - 'what-if' scenarios
  - generic solutions or ...
  - specialist advice
- decision making
  - refining alternatives
  - seeking advice
  - eliminating concerns
  - objections and barriers to the sales process

However, the process is heuristic rather than prescriptive, and the interfaces must reflect differing customer and staff behaviours at each stage of the buying process or reflect differing characteristics associated with individual types of financial product. To address this, a team is established to analyse behavioural issues and produce software application prototypes to experiment with the 'look and feel' of alternative kiosk user interfaces. Ease of use or usability is a key factor in behaviour and an iterative approach to usability needs to be planned for implementation. Further information on usability can be found in [Hutt, 1993(b)].

'Look and feel' prototypes should be developed as early as possible following the business and technology workshops.

2.5 Implementation

Implementation consists of the elements shown in figure 2. Typically, *selecting and verifying technology* is the major element at this stage. It covers detailed functional specifications, hardware and infrastructure software procurement, and the development of the kiosk application to populate the technical and application architectures. Integration of personal computer hardware and infrastructure software components to verify compatibility is an activity frequently overlooked as 'plug and play' is all too often assumed. This implementation includes full systems validation where the kiosk housing comes together with the kiosk application in a test environment.

Engineering processes and information focuses on changes brought about by the introduction of kiosks. Typically these involve kiosk servicing; the replenishment of consumables and routine adjustments, maintenance; repair of failed technology, and other operational procedures. Some systems, particularly those involving monetary transactions, will require operational processes for resilience and data integrity that need to be tested during implementation. Processes for validation of changes during deployment also need to be produced and verified during implementation.

Whenever a new business or information system such as a kiosk is introduced, it inevitably means the organisation has to change in some way – even if it just means extra roles or responsibilities within an existing structure. *Preparing the organisation* is an essential step to ensuring successful deployment covering announcements and briefings to staff, job descriptions, and service, training and documentation plans. Inappropriate staff attitude to kiosks is often a cause of project failure.

2.6 Deployment

Introduction and deployment of kiosk systems are usually in two stages; the pilot stage followed by volume installation. The pilot is typically confined to a small geographic area and a limited number of installations to test public reaction and customers' usage. *Operational control* is concerned with managing and monitoring the kiosks once they are installed. *Operational control* processes defined during implementation are validated during the pilot stage. *Planning and reporting* provides the necessary feedback and statistics to determine achievement of project success criteria. To ensure that appropriate reporting information will be available, the measurements needed must be determined during design.

2.7 Marketing

Marketing plays a vital role in the success of any kiosk project. Most important at the direction setting stage is the understanding of the financial products' buying processes and the anticipated customer interactions. As discussed, these are tested initially at 'customer behaviour analysis'. Products and services offered need to be understood particularly in terms of the interaction time with the kiosk: are they fast transactions or information browse? This reflects on the role of kiosk and queuing patterns. Products and services offered need to be complementary; not transaction intensive mixed with information 'browse' intensive.

In principle, a kiosk may play two fundamentally different roles; as a transaction kiosk and an information kiosk. An example of a transaction kiosk might be for the purchase of travel insurance where information provision is low and the buying process is completed in one relatively short interaction. An information kiosk enabling mortgage sales, for example, may provide information on the whole home buying process as a service to the customers where interaction times are longer and may be repeated for the different stages of the process.

Kiosks should not be thought of as exclusively 'self-service'. They provide an excellent sales tool for assisted or semi-assisted operation, either by branch staff or through video conferencing, ensuring customers obtain the information sought, enabling customer dialogue and providing improved levels of service. There is a major opportunity for development of information kiosks as the front end to existing sales systems in the branch utilising common applications components and multimedia presentation formats. Together, these appear as a seamless service promoting the image of the enterprise and potentially reducing development expenditure [Haynes et al, 1994]. This approach has the potential to provide highly qualified customers to the bank's financial advisors.

2.7.1 Marketing phases

Using the Methods for change model, see figure 2, the marketing activities and contributions at the various phases are summarised below:

- direction setting
  - target markets
  - demographics
  - product selection
- design
  - corporate identity
  - branding
  - measurement requirements
- implementation
  - selection of pilot locations
  - test and control groups
  - internal promotion
- deployment
  - external promotion
  - measurement analysis

3. Building a Financial Kiosk: the Physical Design Process

The nature of the services being offered on the kiosk and therefore, the length and nature of the interaction with the customer, means the physical kiosk design and ergonomics has to support and encourage customer use. The kiosk needs to be styled sympathetically to offer a high quality service while being consistent with the environment and branch architecture. Branding and signage[1], supported by an attractive application to encourage customers' use of

---

1 'Signage' embraces everything concerned with the text, design, and positioning of signs inside or outside the branch.

the kiosk, are key requirements. Following the business planning and design workshops, the environmental aspects of the kiosk enter the design and implementation phases.

The design stage translates the requirements into a visual concept of the environment with design sketches and three dimensional models. The implementation stage covers detailed mechanical design with CAD/CAM automation and the delivery of prototypes. Production models follow at deployment.

Virtual reality techniques can assist with the view and placement of the kiosk in the branch, particularly if the kiosk is being introduced during a complete branch redesign. In any case, each branch has to be surveyed and designed to accommodate kiosk systems.

3.1 Kiosk Design

The kiosk environment includes the housing of the information technology and its surrounding environment - both equally important to the business application and services being offered. In parallel with the definition of the application and technology architectures, the physical environment needs to be defined and developed. Key considerations are security, privacy and, particularly where video conferencing is employed, lighting aspects such as glare and reflections. Sunlight is a particular consideration and VDU automatic brightness compensation controls such as the ICL ErgoPro 'Autobright' feature becomes an important design consideration.

ICL has developed a method for the physical aspects of kiosk design which is described below and has been proven through the development of a number of kiosks.

3.1.1 Design

The output requirements from the technology workshop are taken by the concept designers and used to produce design sketches for discussion and review. Examples of concept design sketches are shown in figures 4 and 5. After refinement of the concept design and initial design drawings addressing spatial requirements of components, accurate full-size models in polystyrene foam are produced to test the design concept, ergonomics and serviceability. These models, as well as providing visualisation of the design, are an essential first step in analysing usability such as reach and movement; particularly important for the touch screen and peripheral use such as a magnetic card reader and PIN pad.

The next step is to produce semi-working wooden models using the selected peripherals and components. From a design point of view, this confirms spatial requirements and allows cable routing to be assessed whilst providing the next iteration of functionality and usability verification.

3.1.2 Implementation

This stage involves the detailed kiosk development of component assemblies leading to prototype manufacture. The outputs from this phase are detailed mechanical, electrical and electronic design drawings, specifications for prototype manufacture including numerical control program and plastics tooling as appropriate to this stage. As design drawings are produced, CAD/CAM modelling takes place to present the kiosk in its intended finished form. An example of CAD/CAM rendering (computer graphic views of the system) is shown in figure 6. CAD/CAM modelling is invaluable in order to refine the design, verify colour use, styling and finishes, and, potentially, positioning in proposed environments.

Figure 4 Concept design sketch – standing format

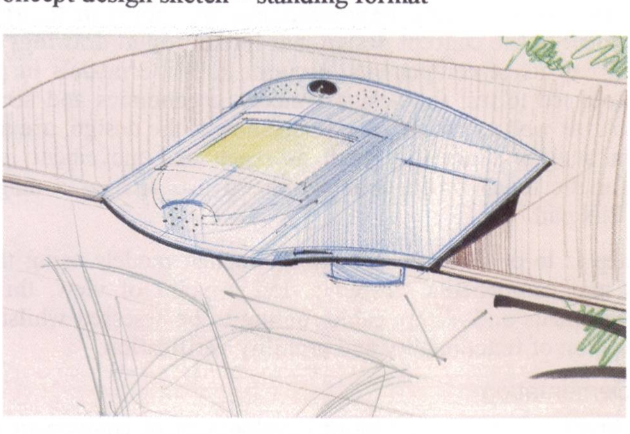


Figure 5 Concept design sketch – seated model

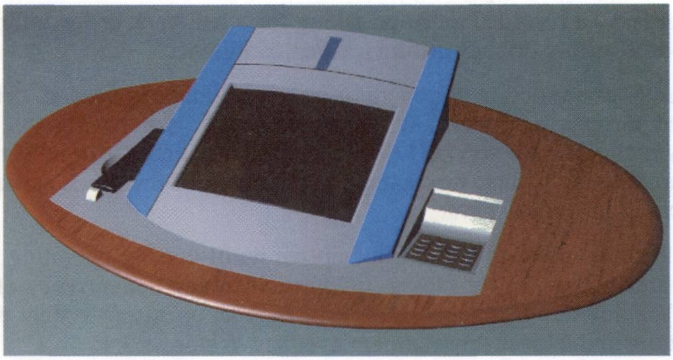


Figure 6 CAD/CAM rendering

Prototype kiosks produced at this stage come together with the application for final systems testing and usability assessment.

Safety approval certification, radio frequency emission verification and other regulatory approvals can begin with the prototypes produced at this stage together with a serviceability assessment to ensure maintenance requirements are met.

4. What's in a Typical Kiosk System?

The heart of the system is typically a 66Mz Intel 486 or Pentium-based PC with 32 Mbytes of memory and 1 to 2 gigabytes of hard disk depending on the requirements for storage of video clips. The CPU can be enhanced with a graphics accelerator card ensuring smooth running of video clips and sub-second response times on animation. A touch screen, laser or receipt printer and speakers are incorporated within the kiosk housing. The size of visual display unit screen is an important consideration and is driven by the volume and complexity of information to be displayed. A transaction kiosk may require a 14 inch display whereas an information kiosk may need a 17 inch display. Optionally, a 'telephone' handset may be provided to allow customers to listen privately to sound output by switching from speakers to handset as it is raised. The handset is also an added feature to assist people with hearing difficulties. If video conferencing is provided, a video camera is mounted within the kiosk housing. For video conferencing, the handset becomes essential from the usability 'level of comfort' aspect and so as to have a private conversation with a financial adviser. Other peripherals include magnetic or smart card reader and a PIN pad.

Kiosks may contain various in-built sensors detecting the proximity of a customer, the presence of printout in the output paper tray, the telephone handset raised and the service key activated. The proximity detector can be used by the application software to initiate a welcome or "attractor" sequence and "paper detected in the output tray" is used to prompt the customer to collect it. Detection of the "handset raised" is designed to allow customers, for example, to leave a recorded message asking for a call back or for information to be sent or to ensure privacy during a video conference.

With any self-service device, an operator or service mode is needed. This can be provided through a standard PC keyboard which is not accessible to the public. By inclusion of a service key switch, operator action can be simplified by a pop-up menu allowing, for example, calibration of the touch screen or to facilitate close down of the system.

Kiosk applications demand a full pre-emptive multi-tasking operating system to ensure smooth running of moving picture video and concurrency of tasks. Microsoft Windows 95 is a good example of such an operating system offering major benefits in the kiosk environment over say Windows 3.1.

5. Personal Computing Versus Self-Service

Most personal computing products are not intrinsically designed for use in the self-service environment; they assume a local user familiar with the system. Typically, any error or warning situation causes a 'pop-up' window to inform the user of the condition. In a self-service environment, the kiosk user has to be protected from any such messages that the PC system software might display. Additionally, the user interfaces to many personal computer applications assume the use of 'pull down' menus and mouse operations that are incompatible with self-service, thus 'off-the-shelf' applications cannot be easily utilised.

Peripheral management is another area not addressed by personal computer technology. A typical printer for example, uses its own display panel to provide information on its status such as, paper out, toner low and so on while passing only 'OK' or 'error' status to the PC. The availability of more comprehensive peripheral status information is essential to the management and servicing of kiosks. Enhanced printer interfaces and management sub-system products that address this issue are now beginning to emerge such as Markvision from Lexmark. Self-service peripheral management requirements are now being recognised and standardised as part of the work undertaken by the Banking Solutions Vendor Council (BSVC) in its extensions to WOSA (Windows Open Services Architecture) for the financial services industry, see [BSVC, 1995].

6. Self-Service Based Video Conferencing

Video conferencing is a powerful technology for kiosk systems, facilitating access to experts or advisors potentially around the clock and is one of the key requirements. Most personal computer-based video conferencing products are designed as 'person-to-person' videophones assuming a competent user of the PC system. When video conferencing is moved into a self-service kiosk environment, it is typically used by a customer to obtain information on products or services from an advisor based at an organisation's call centre. This environment places two new demands on the video conferencing system:

- a simple customer interface
- a trace of what the customer has already done.

The kiosk customer interface has to avoid the screen-based call set up processes normally associated with making a PC video conference call. This needs to be replaced with a simple, single touch of a 'talk to someone' button on the screen. When a call is made, there needs to be an indication available to a call centre advisor of any self selection of products or information searches already carried out by the customer.

To achieve the above requirements, ICL's TeamVISION video conferencing system has been enhanced with a full Applications Program Interface, API. This allows a kiosk application program to take control of the complete video conference, making all outgoing call connections on behalf of the kiosk user. The kiosk application itself needs to record the trace of the customer dialogue but rapid transfer of this information to the call centre is facilitated by TeamVISION.

Another powerful facility in the self-service context is screen-sharing, that is to say the ability for one PC to take over control of another. This can be used during the video conference dialogue for the advisor to take control of the kiosk in order to point out details or to assist data entry on the customer's behalf – particularly when typing information on to screen-based forms! A remote printing facility allowing the call centre advisor to initiate printing for the customer at the kiosk is also available.

With TeamVISION, audio telephony facilities operate normally and voice only telephone-based services could be offered too at the touch of the screen, for example, to call a credit card hot-line.

Video conferencing demands the implementation of a *call centre* which is a major project in its own right. The techniques of audio telephony call centres are now fairly well understood. Video extends the requirement for example to include call centre lighting and background, presentation of corporate image in terms of advisors' dress simulating that of the branch, complementary scripting systems – usually the domain of Computer/Telephony Integration systems which are currently unable to handle video conference calls, 'video phone' operator techniques such as body movement and facing the camera.

7. Software Infrastructure

Figure 7 shows the typical software infrastructure for a kiosk system with emphasis on low-level aspects. Other examples may include on-line access to central systems. This example shows the business application, middleware, of which video conferencing is a special case, peripheral drivers and systems management. A business application can take many forms but will usually need to be supported by middleware which handles the intimate dialogue with peripherals and simplifies the interface to the business application. The middleware printer manager, for example, supports the additional demands made on printers in a self-service environment as well as providing a high level forms interface to the business application. A forms definition tool allows forms to be predefined and held by the printer manager. The business application then simply adds any variable data.

Operations monitor communicates warning conditions to systems management and collects statistics on peripheral usage to assist in preventative maintenance, for example, when a given number of reads have been carried out by a motorised magnetic card reader. Warning conditions typically translate into events that are automatically notified to centrally located operations personnel.

The journal, sometimes called the kiosk 'black-box', monitors all usage of the business application, logging navigation paths through various screens, the actions taken within a screen and the time taken to move between screens or to perform an action. Journal analysis provides the insight into system usage and gives invaluable information for both marketing and usability assessment.

Systems management, as well as being concerned with handling warning conditions, also covers on-line distribution and installation of new software and remote access for systems administration. A number of products are available to address systems management in the areas of event management, software distribution and installation, and remote administration.

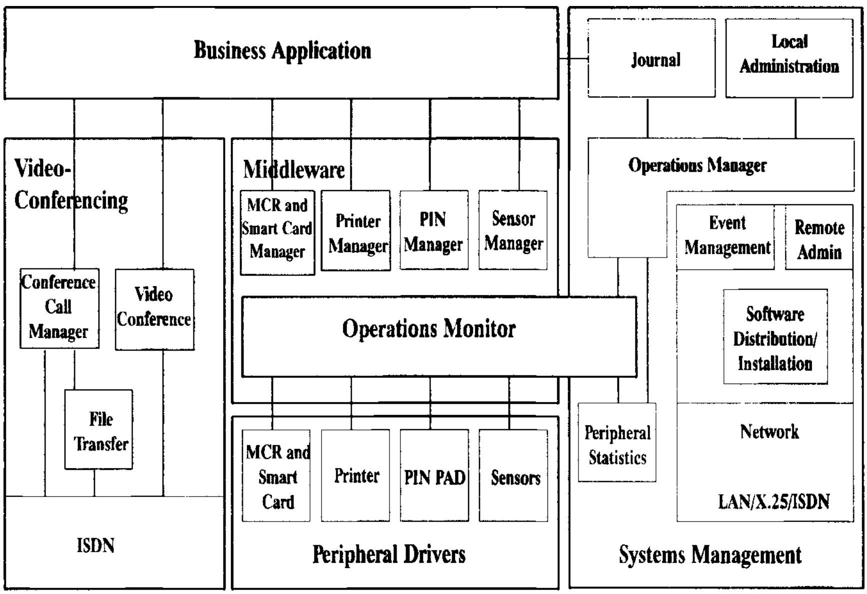


Figure 7 Kiosk infrastructure software

8. Ten Success Factors for a Kiosk Project

Kiosk projects need to succeed in all of the following areas to fulfil their potential; failure in any one of these could result in failure of the whole project.

- marketing – products and promotion
- usability
- staff acceptance
- environment
- regulatory/security
- business application software
- technology infrastructure
- systems management/support processes
- call centre
- partnerships

The above points have been covered already with the exception of partnerships. It is unlikely that any enterprise today has all the necessary kiosk skills in-house and available, thus partnerships with vendors represents a potential way forward.

9. Forward Projection

The following is a view of how kiosk systems will develop:

1-2 years	Increasing numbers of information and transaction kiosk pilots. Early adopters moving into volume installations.
3-5 years	Information kiosks integrated with branch based multimedia advisor and sales systems. Installation of kiosk based 'virtual' shopping malls offering banking, retailing and other services.
5-10 years	General acceptance of kiosks as alternative delivery channel. Technology advances such as flat panel displays, voice input influencing design and capability.

Initially, information and sales kiosks will be implemented stand-alone rather than fully integrated into the branch system and typically supporting the 'information seeking' aspect of the buying process. Over time kiosks will have a major impact on the sales process and the next generation of sales advisor applications will be multimedia-based, tightly coupled to the kiosk and sharing a common application. The continued cost/performance improvements of multimedia PCs will accelerate the installation and deployment of kiosks.

Smart card technology advances are expected to have a major impact on kiosk applications. The smart card provides an ideal vehicle for a customer to capture and retain personal information derived from interactions with the kiosk. The smart card can be used in conjunction with customer loyalty schemes and services promotion through the kiosk.

The personal computer population in UK homes is currently about 12% rising to a predicted 49% by the millennium. Cable technology over the next few years will deliver higher bandwidth to more and more homes passing an expected 71% of homes by the end of the decade. While education and entertainment are two of the key drivers for personal computer and cable technologies, they facilitate on-line access to information, harnessing the capabilities of multimedia and self-service to deliver the 'genuine' armchair home shopping experience: the home kiosk.

10. Conclusions

Kiosks offer a means for controlling costs, upgrading the role and effectiveness of branch staff and, most importantly, meeting the needs of customers whilst maintaining an ever improving level of perceived customer service.

Multimedia kiosks signal a significant change in the way a branch of the future will talk to its customers. Traditional glass screens and counters are being replaced by an open-plan informality, as staff step out from behind the barricades to sit down side by side with the customer.

Kiosk and the self-service environment demand specifically designed personal computer technology particularly in the areas of user interface and systems management. Kiosks, however, should not be thought of as technology projects, more the provision of a tool to assist in the generation of business and the development of customer relationships.

Compared with ATMs, which took ten years to become fully accepted, the future for interactive multimedia kiosks is only just around the corner.

References

BRUNT, R. and HUTT, A.T.F. The Systems Architecture: an introduction, Cambridge, Prentice Hall, 1992. (ISBN 0-13-560186-X)

BSVC (Banking Solutions Vendor Council), WOSA Extensions for Financial Services, Applications Programming Interface and Service Provider Interface Specification, February, 1995

COMBRIDGE, M. Quality function deployment, ICL, 1992

HAYNES, M.W., IBBETT, G. and WALKER, D.J. Exploiting Client-server Computing to Meet the Needs of Retail Banking Organisations, *ICL Tech.J.* Vol.9 Iss. 1 pp.47-66, 1994

HUTT, A.T.F., (a) OPEN*framework* Methods for change, ICL, 1993

HUTT, A.T.F., (b) Usability, Cambridge, Prentice Hall, 1993.
(ISBN 0-13-630930-5)

HUTT, A.T.F., (c) OPEN*framework* Developing architecture, 1993

Biography

Mike Shaw has worked for ICL since 1967 with a career spanning various roles including hardware design, operating system software development, project management, marketing, consultancy and business management. He was part of the original team that created OPEN*framework*.

He is currently the manager of the Major Projects group within ICL Kiosk Systems. He has been involved with a number of multimedia kiosk projects of which the most recent experience has been the development of Interact with the Nationwide Building Society, UK in the role of ICL project manager.

High Availability Manager[1]

D. Stewart

ICL Volume Products Division, Kidsgrove, UK.

Abstract

This paper describes some of the significant events in the development of High Availability Manager (HAM), from the summer of 1993, through to the product in the field today in mid 1995. The history and background to the development, the people who designed the code and the processes used to control the development will be described in outline. The HAM product, together with the challenges faced and overcome during the development and potential future HA products are next described. The paper ends with a look at potential future developments.

1. Introduction

The paper outlines the history and current status of High Availability Manager, an ICL product designed to ensure a high level of availability without the complication of a totally fault tolerant system.

1.1 The Nature of "High Availability"

It is to be noted that High Availability (HA) systems are more primitive than full fault tolerant systems. Whereas in full fault tolerant systems all errors will be managed in such a way that the end user sees no impact at all from the failure of a hardware component, in an HA system, the failure of some components will result in a short break in service. This break will typically take a few minutes before the application is available for use. Full fault tolerant systems, sometimes called continuous processing systems, raise different cost of ownership issues from those in HA systems. In administration, for example, continuous processing systems with replicated hardware (sometimes three or four times replicated) are presented as single systems and are managed as single systems. This takes less administrative effort than an HA system where two separate systems are seen by the administrator. On the other hand, in an HA configuration, both of these

---

[1] This paper was a prize winner in the OPTA competition at the 1994 ICL Engineering Conference.

separate systems will, in normal circumstances, do useful work for the users. By contrast in continuous processing systems all the hardware, including redundant hardware, is dedicated to a single workload.

Sometimes, applications written to take advantage of the continuous processing for fault tolerant systems, require to conform to particular constraints. In HA systems, applications are exactly the same as in an equivalent single system.

Thus it can be seen that there are advantages and disadvantages in both approaches to higher availability.

The basic mechanisms used by ICL to achieve High Availability are simple. All hardware is replicated. A software application monitors the state of all hardware and in the case of a minor hardware error, reports the error for action at an appropriate time, or for serious hardware failures, switches any mission-critical applications to a working system. For user data, replication of the disk hardware on its own, is not enough to ensure the ability to continue to run an application. If a disk fails the data on the disk needs to be protected using a reliable method, such as, disk mirroring (RAID level 1), or other higher level RAID technique. Using one of these methods has the added advantage that, in the event of the failure of an application data disk, the system will continue with no immediate break in service to the end user. At some time it must be arranged that the data is brought back to a protected state using another disk. This can be managed in a number of ways which will be described later in this paper.

As some aspects of continuous processing systems are superior to HA systems, the question, "Why buy HA instead of a continuous processing system"? must be answered. HA provides a superior combination of cost and flexibility for many applications. Continuous processing systems need all of the underlying complexity to be present all of the time. HA systems make use of replicated standard systems, which in the case of small-to-medium size servers, are becoming low-cost commodity items. This situation is reminiscent of what happened in the PC market a few years ago.

An internal cost comparison by ICL in 1993 showed that an HA solution could cost between one quarter and one fifth that of a full fault tolerant system, for a system of equivalent processing power. The cost performance trend for Industry Standard Server products has remained very favourable to customers during the period since then. An additional advantage with High Availability systems is that the customer can choose how many of the potential availability components they wish to buy; one example is an Uninterruptable Power Supply (UPS). The customer might decide that he has a low risk of failure of the mains supply or, as in hospitals, they may already have auxiliary generator capability. Another example is to have a dual LAN, this may offer availability advantages, and it is supported in HAM. But in the end, the IT provider must supply what customers feel is right for their situation, and what they are willing to pay for.

Industry watchers consider that the advantages of fault tolerant or continuous processing systems are such that there will continue to be a level of demand for them, but the lower cost and flexibility of HA in being able to meet competitively a wide range of availability requirements, will ensure that HA will meet the needs of the majority of customers.

1.2 Influences on the ICL Product Development

Quite reasonably, our sponsor and first customer made it a condition that the product produced would become a standard part of the ICL product line. For this reason, the development has used the ICL standard development processes.

According to D.H. Brown Associates, [D.H. Brown Associates, 1994] in a paper called "UNIX Leapfrogs in Commercial Availability" the first Commercial UNIX HA capability came from Hewlett Packard in 1989 with a product called SwitchOver UX. A number of years passed before the next set of UNIX HA products emerged from OpenVision, IBM, Data General and NCR in 1992. ICL cannot claim to be among the leaders in this technology as our first product appeared in December 1993. While ICL was not among the first set of HA product announcements, there were some compensations for later entry into the market as will be described later.

The first HA product which came to notice was IBM's HA 6000 which runs on RS 6000 under the AIX Operating System. AIX is the Advanced Interactive System, IBM's own version of the UNIX Operating System. This product was, and is still, very well promoted by IBM. Even in 1992 it was an obvious candidate for an equivalent, preferably a better product, from ICL.

Since that time we have collected a large dossier of information on competitive HA products which, in varying measure, provide standby systems for mission critical applications. The main systems we have met in direct competition are HA 6000 and HACMP from IBM, Lifekeeper from NCR, Reliant on Pyramid, and Open HA from OpenVision running on SUN.

1.3 The ICL Approach

At the start of the High Availability Manager development ICL was in the favourable position of having the agreement from its technology design group that the design proposal was technically feasible and also had a definite customer. The advantage of having a customer with specific needs and timescales also helped to distil the requirements to a realistic set providing the required functions without over elaboration.

Proprietary mainframe environments have had standby capability for years. In many ways, ICL's Series 39 Mainframes are still unsurpassed in the facilities they offer in this area, through the combined capabilities of the VME Operating System and the Macrolan Fibre Optic interconnect technology. In a manner similar to proprietary mainframes, early High Availability UNIX products were constructed out of proprietary, add-on facilities. These pioneer products did provide a useful capability, but the degree to which they could be described as "open" can be seriously questioned.

Several conflicting goals had to be balanced in the ICL development programme. These included the very tight timescale to which we agreed to produce the initial product, the high quality goals set to ensure that the High

Availability product would indeed help to improve availability and our desire to use standard, modern, open components.

The Programme and Product Plans were set after asking advice from a wide cross section of technical and commercial people in ICL. The compromise underlying these plans was not without risk. We chose an all modern software set, some components of which were not generally available at the start of our programme, and planned more time for validation than the normal ICL development planning template required.

The HAM product was developed by ICL using Industry Standard components for both hardware and software, ensuring the maximum degree of openness. Only the very heart of the monitor/automatic operator facilities of the High Availability Manager application was written as bespoke new code.

All of the initial developments in High Availability Manager focused on systems which ran under the UNIX Operating System. Since then HAM has been further developed and adapted to provide a greater range in functionality on UNIX systems and the original V1 product has been extended to provide an HA platform for Servers running NT Advanced Server, Operating System.

2. The ICL HA Manager Product

Figure 1 illustrates a typical HA system where a mission-critical workload can be run on one of the processors and another useful, but not critical, workload can run on the other system.

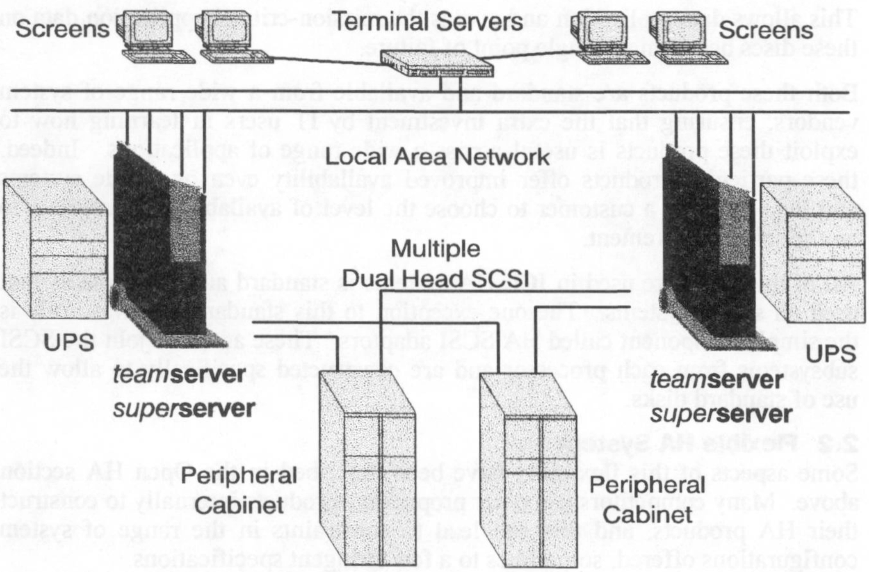


Figure 1 A typical high-availability system

In addition to the basic resilience capability, the ICL HA solution provides a number of valuable additional attributes. These include a greater degree of openness and flexibility than alternative HA solutions through the use of open, industry standard components. Also there are features which make the HA system easy to set up and, most importantly, a wealth of *resilience-enhancing* features.

2.1 Open HA Systems

By coming to the market later, ICL could integrate its product using industry standard components. Specifically the Veritas software products, VxFS and VxVM, offer a range of advantages as described below.

The standard software components used to develop HA Manager were:-

UNIX V7, Operating System.
VxFS, Veritas Filestore.
VxVM, Veritas Mirroring SW.

All these components were new at the time of the original development and are still used in current UNIX products, although there are more choices of components in the current products as will be described later.

VxFS offers fast filestore recovery by maintaining a non volatile "intent log" which in the event of a system break indicates the relatively few files which have to be checked for consistency. This offers a massive speed improvement over the older standard UFS, which could make no assumptions and was forced to check the entire filestore using the check utility called FSCK.

VxVM offers disc mirrors and has an easy-to-use X Windows interface VxVA. This allows data replication and avoids the mission-critical application data on these discs becoming a single point of failure.

Both these products are standard and available from a wide range of system vendors, ensuring that the extra investment by IT users in learning how to exploit these products is useful across a wide range of applications. Indeed, these particular products offer improved availability even on single systems and help to allow a customer to choose the level of availability appropriate to any specific requirement.

All of the hardware used in ICL HA systems is standard and the same as that used on single systems. The one exception to this standard hardware rule is the simple component called HA SCSI adaptors. These adaptors join the SCSI subsystems from each processor and are constructed specifically to allow the use of standard disks.

2.2 Flexible HA System

Some aspects of this flexibility have been described in the Open HA section above. Many competitors use older proprietary products internally to construct their HA products, and this can lead to constraints in the range of system configurations offered, sometimes to a few stringent specifications.

Additional flexibility was introduced in HA manager Version 2, where extra interfaces are provided to allow HA Manager to monitor application components as well as hardware units and to allow flexible action should an application fail.

2.3 Easy HA System

In all HA systems the *administrator* must describe the system hardware configuration. HA Manager offers the administrator a full set of configuration data ready to use. To chose the particular configuration set for any specific system, the only operation required by the administrator is a very simple edit. This feature goes some way in reducing the management advantages of continuous processing systems, described earlier.

2.4 Robust HA System

HA Manager offers several unique features to help ensure that the ICL system provides the required functionality. The HA manager application components in each of the live and standby processors are individually robust, in the sense that the main controlling processes detect the failure of another controlling process and will recover by starting a new process when required.

HA Manager components in each system run tests on all major hardware components, even those not in immediate use by applications. In addition, the HA Manager in each system communicates with its counterpart in the other system to check the state of the other system. This is the primary mechanism which allows detection of a failure of the other system and makes the controlled takeover of a mission-critical application possible. For faults which are less drastic, the two systems compare their views of the error before reporting. This improvement in the diagnostic capability gained by comparing notes from both systems is an area where the ICL solution has been patented.

In addition to improving fault detection, HA Manager provides fault reporting. All faults are logged by high availability manager in each system and to ensure that both sets of logs make sense when later compared, HA manager ensures that the clocks in both systems are synchronised within a fixed tolerance.

In the event of controlled takeover as briefly described above, it could of course be catastrophic were this to occur when the original live system was still functioning. This has been termed "false partitioning" and must be avoided at all costs. The HA Manager solution is to have multiple communication routes between the systems. The primary communications are via LAN but alternative communication paths via shared discs are also used and are constantly monitored. The LANtalk and DISCtalk approach in HA Manager is, as far as we know, also an ICL unique feature. Only when all the routes (at least three) cannot communicate is the decision made to "*failover*" to the standby system. (The term "failover" is widely used in High Availability products to describe the act of one processor taking on the application of a processor which has failed).

HA Manager is implemented as an application, with no dependency on the inner workings of the operating system. This was a deliberate design choice. The advantages are that provided HA manager is given the correct priority level it ensures that applications are able to be run. This contrasts with alternative designs, notably that adopted by IBM where the HA component is implemented in the AIX kernel. With the HA component in kernel, it is theoretically possible for the HA communication to continue as normal while the real application system which the users need may not be available. In this respect HA Manager also checks the functioning of the operating system base. Another advantage in having the HA manager as an application is that it is much easier to port an application of this type between systems and even to alternative operating systems.

2.5 Maintainable HA Systems

HA systems of course have to cope with hardware failures, but in the rare event of such failures the systems have to be maintainable. This can be straightforward in some cases such as a serious hardware failure where the failing processor is isolated and can be repaired off line and re introduced when corrected. A more difficult choice arises when a disk plex is lost. As previously stated, the system can be allowed to continue to run for a time determined by the operating policy and the priority of the current work. This continuation however, cannot be allowed to run indefinitely in the state where only one copy of the mission critical application data is being maintained.

A number of options to help the operations policy are provided. The simplest of these is simply to allow the work to complete to a point convenient to the operation and stop the system to allow the removal and introduction of an alternative disk. This disk can then be brought into full resilience using the on line facilities of VxVM without further interruption to the system. Alternatively, if the system is fitted with 'Hot Pull' capability, this operation can be carried out without the need to wait until the job is complete. At the time of writing only a subset of ICL Servers can take advantage of Hot Pull facilities. An even more sophisticated solution is available for SPARC SuperServers, this is called Hot Spare. In this case the software detects the loss of a disk and allocates a new disk from a previously defined list of spare disks and automatically rebuilds the mirrors.

3. Technical Challenges

Many complex challenges were encountered and overcome in the production of HAM as a new product.

They ranged from the complexity of ensuring that in the event of failover the standby system assumed the full identity in all respects of the previous system, to working with very tight deadlines with dependencies on system components from a variety of sources. The final challenge was to project the product appropriately to customers.

The technical challenges faced by the development team were, in the main, the most controllable. Like any other major product the quality "Potential for Change" had to be built into the design principles from the outset. The development required a first rate designer to react with speed and accuracy to the problems which were encountered. This prescription sounds simple and it was, for this team, because they had Richard Taylor to do the hard bits.

The programme dependencies were more problematic; delays from suppliers of some software components had to be managed. Sometimes slips in deliveries were reported very late. Fortunately, when these products did arrive they worked well and the programme survived. Other learning points in the programme area were concerned with hardware deliveries. ICL's HA system needed a single very simple non-standard hardware component to join the IO systems, and to block unwanted hardware resets between these systems. Managing the timely delivery of these components of our HA offering turned out to be difficult for people with previous experience only in software development. This team now understand more about hardware development processes and are in better shape to manage the total systems aspects of projects than they were at the outset.

4. Marketing Challenges

Management of customer perception has also proved difficult at times. The differences between HA and continuous processing capabilities are real and important. The relative subtleties of the difference may not always have been easy to understand. Regrettably, one customer bought HA Manager in the belief that it was a full continuous processing system. ICL makes every endeavour to train its staff to describe its products accurately and maintains a high degree of vigilance that everyone who has responsibility for this communication understand the products. This particular problem as described above occurred overseas, where language difficulties may have been the root cause of the communication problem.

Another aspect of the difficulty of describing product capability accurately, at any stage, arises from the fact that products are developed and improved continually. Sales channel staff and customers want assurance that they are buying a product which has an ongoing future. They want to see a forward vision and the developments proposed to underpin this vision. As described in the section Future Prospects below, ICL has just such a forward view for HA products. A number of facilities, not present in the current released product, are already in development. Thus it is necessary to be clear which features exist in the current products and which are under development for incorporation in future products.

5. Field Experience

HA Manager is one of the Server Systems products collectively known as ManageWare. The products which we offer within ManageWare are added-value products which enhance the capabilities of operating systems. These include backup and restore, print management, transaction processing, availability, resilience and performance enhancing products some of which are mentioned here as component parts of HA.

HA Manager was released in December 1993; since then ICL has sold more than one hundred systems.

As already stated, HA Manager uses standard hardware and software components and this has simplified field maintenance through not needing to stock specific different components as replacements, such as twin tailed disk types needed for some HA implementations. Customers benefit in two ways from the use of standard components; the industry standard components will always be lower in cost than components which are more specialised and sell in lower volume; again the skills needed to operate the software are more widely available in the market for industry standard software components.

High Availability has proved to be a key to satisfying a number of specific business requirements. These include command and control systems, fast response warehousing, telephone billing, financial and retail systems.

6. Future Prospects

The initial product in December 1993 was HA Manager V1. This was a simple standby only system, where the standby processor did no useful work.

The second release was HA Manager V2 in July 1994: this added the capability to run non-mission-critical workloads in the standby processor and added application monitoring capability.

HA Manager V3 has been just been released (in mid 1995). This adds support for the UNIXWARE operating system and faster failover.

Looking further ahead, the development team is considering a wide variety of offerings. These include:-

6.1 Dual Mission Critical

This allows two processors both running mission-critical applications to provide standby capability for each other.

6.2 Hardware RAID 5 Support

This allows the option of using either the current software mirror technique or RAID 5 discs to protect the customers' data. The current product uses software RAID 1 (mirrors) which is both high performance and cost effective.

6.3 Disaster Standby

Current HA systems require the mission critical system and the standby system to be in close physical proximity. The requirements of disaster standby are that these systems can be separated, typically by a distance of a few kilometers.

6.4 Availability Clusters

The current HAM system is aimed at two processor target configuration where one processor acts as a standby for the other. A more flexible approach where an application may be executed efficiently on any one of a much larger set of possible processors will be enabled by new industry standard hardware developments. The hardware developments which will enable this new HA architecture includes Fibre Channel Arbitrated Loop (FC AL) and SSA architecture disks. HA products which will exploit this technology will be developed in line with availability and customer demand.

6.5 Multinode Clusters

This is a true distributed processing environment where any processor in a cluster can act on behalf of any other. This is a useful means of providing a high-power system in a cost-effective manner, but in the longer term such processing will be carried out on massively parallel systems of the type described as MegaServers (GOLDRUSH). [Watson & Catlow, 1995] [Watson & Robinson, 1995].

6.6 Ease of Use

As described earlier in section 2.3, HAM systems already provide some ease of use features to assist administrators. In line with modern ease of use requirements, a GUI interface will be provided to make administration simpler still.

7. Acknowledgements

Almost all the development and testing of High Availability Manager was done by Richard Taylor and Paul Speed. Besides writing much of the HAM code, Richard Taylor designed the product, wrote the Product Plan and most of the text for the manuals.

Starting two months into the development, Paul Speed questioned the intended structure of the product, which in turn led to a better specification and implementation. He independently developed the HAM sub-tests which constantly check the availability of all major components in a HA System. Paul was co-developer with Richard Taylor of the TCP/IP failover and shares the patent with Richard for "Auxiliary Root Technology" which maintains a constant machine identity in the case of failover. This is one of two patents which have been applied for in the course of the development of HAM.

Dave Stewart, was manager and main external promoter of the product, in its initial development stage.

Without the active help of the following people, this product would not have been possible.

Paul Townsend, Barry Read, Steve Hilditch, Alasdair Velzian, Janet Cropper, John Lai, Mike Martin and Frazer Campbell

References

D.H. BROWN Associates, UNIX matures in Commercial Availability, private report to PTEP sponsors, referred to by permission of D.H. Brown Associates Inc., Port Chester, NY 10573 USA, March 1994.

WATSON, P., and CATLOW, G. *Ingenuity*, 10(2), pp. 191-205, 1995.

WATSON, P., and ROBINSON, E. *Ingenuity*, 10(2), pp. 206-219, 1995.

Biography

David Stewart

Dave Stewart graduated with a BSc in Physics from Strathclyde University in 1968 and took a circuitous route into the computing field via writing Computer Aided Design (CAD) tools for Optical Systems. He joined ICL in 1970 to work on CAD tools for Computers. During the last twenty five years he has carried out a variety of Technical and Management functions working in diverse areas such as I/O Subsystems within VME Operating System, Test Software, Communications, Database and Transaction Processing. From 1988 to 1990 Dave was the Manager of VME Technology which was the main technology team leading developments in ICL's VME operating system, IDMSX Database and TPMS Transaction Processing Monitor. In 1994 he led a small team which developed ICL's High Availability Manager, a product which won the internal ICL competition for the Outstanding Performance in Technology Award.

The Virgin Global Challenger

Simon Forse

Chief Engineer, Lindstrand Balloons, Oswestry, UK

Abstract

Richard Branson the British entrepreneur is sponsoring the design and construction of a specialised balloon for a round-the-world flight. He himself will travel with the designer Per Lindstrand and one other crew member. The requirements for and the actual design of the balloon and its capsule are outlined, together with the design principles first proposed by Rozier over 200 years ago. The balloon will travel in the jetstream, easterly winds in the upper atmosphere above 30,000 feet. Since the intention is that the balloonists shall not land during the voyage, the capsule must carry all necessary food and life support facilities. The design draws on experience of previous successful transatlantic and transpacific flights; extensive use of computer-aided design and flight simulation on ICL computers allows many design and operational details to be tested in advance on the ground, though test flights are also planned.

1. Introduction

London: England: 25 July 1995. British entrepreneur Richard Branson and balloon manufacturer Per Lindstrand announce their intention to undertake the last world record left on earth, a non-stop circumnavigation of the world by balloon.

This paper sets the scene on this great adventure and the role British Industry, including ICL, is playing.

Commenting on the project Richard Branson said: "I believe this is the last great aviation record and adventure left on earth. The project is in the spirit of Jules Verne and there would be special significance in bringing the record to Europe before the Millennium. We will attempt to circumnavigate the globe using only the forces of nature – the wind – to propel us around the world in about eighteen days. We appreciate that there are others ahead of us in the race to develop the right type of balloon but we are confident that with the help of British Industry we can catch up. What a wonderful spectacle it would be if we could persuade all the balloonists to take off on the same day and turn it into a race".

Per Lindstrand added: "Richard and I have been through many adventures together but this is the ultimate challenge. We will travel around 20,000 miles [50,000 kms using only the jet stream winds at over 30,000 feet [8 kms] to propel us at speeds up to 250 mph [380 kmph]. This hostile environment will test stamina to the limit but it will also use the latest technology in materials for both the balloon envelope and the capsule in which we will live for over two weeks. Paradoxically, although we believe that the latest technology has made this attempt possible the concept of our Rozier balloon is over 200 years old and our flight has been a dream of people for almost as long".

Richard Branson and Per Lindstrand were the first to cross the Atlantic Ocean and then the Pacific in a hot air balloon, both of these exploits placing them in the Guinness Book of World Records.

For this latest challenge they will be joined in their pressurised capsule by a third crew member, enabling one member of the crew to sleep at all times.

Going around the world by balloon will necessitate sustaining a flight of between eighteen and twenty-one days duration, a substantial test of the team's ingenuity and endurance.

The project has a substantial specialist support team available, with practical experience of two previous record breaking flights, combined with technical skills that cover every aspect needed in the construction of the balloon and capsule and the planning of the flight itself.

ICL has provided the computing equipment used in the design and manufacture of the balloon and the administration and management of the project.

With this attempt Per and Richard will set out to achieve the first non-stop circumnavigation of the world by balloon. This means in effect, a tripling of the longest balloon flight so far which was Per and Richard's flight across the Pacific in 1991.

In order to qualify for a global flight the balloon must cross the launch longitude having covered a distance of 4 radians of the earth which, at the latitude chosen, will be 25,500 kilometres (13,800 nm). The balloon is not allowed to touchdown anywhere or to pick up fuel, gas or ballast during the flight.

To achieve this the balloon must seek out the strongest possible winds which is the jetstream, visioned as bands of fast winds at an altitude of 30,000 ft. The jetstreams do not go uninterrupted around the world and there are areas where the balloon will have to manoeuvre from one to the other using much more slowly moving airstreams. The time the flight will take is entirely up to the wind but based on observing previous wind patterns, an estimate of 14 - 21 days is realistic.

Only one aircraft has so far managed to circumnavigate the world without refuelling when Dick Rutan flew his 'Voyager' for 9 days, around the world in 1986.

The Virgin Flyer which crossed the Atlantic in 1987 and the Pacific Flyer that crossed the Pacific in 1991 were both hot air balloons. This time Per and Richard have chosen a helium balloon which has a much longer endurance and in this case, is capable of staying airborne for up to 21 days. A helium balloon is less manoeuvrable than a hot air balloon and as yet, no helium balloon has sustained flight in a jetstream so before the big balloon takes off in November/December this year, a number of test balloons have been built to explore this and other new aspects of gas balloon flying.

2. The Weather

The weather or rather the aerodynamic forces created by the weather is what propels a balloon. By definition a balloon is an aerostatic vehicle incapable of manoeuvring horizontally by its own force.

The strongest force of weather is a hurricane which needless to say, would tear a balloon apart very quickly and will represent the greatest danger to the entire Global balloon project. The typical areas where hurricanes will occur during our flight period would be in the Philippines and around Florida. Such a hurricane can develop from nothing to full blown strength within 3-4 days i.e. could develop within the time that the balloon is airborne which, most certainly would force the balloon to carry out an emergency landing.

The second strongest force created by the weather is the jetstream which is formed by cold polar air descending south and colliding with moist and warm tropical air. In the interface of these two air masses, the air is squeezed eastwards at a very high speed creating what we call jetstreams. Their width can vary considerably, say from 50-100 miles but they are rather thin, normally only occurring in the tropopause i.e. between 29-37,000 ft. To qualify as a jetstream the wind speed must be a minimum of 50 knots but the average jet is more likely to be 100 knots and over the Pacific, a jet can have speeds well in excess of 200 knots. During the Pacific balloon crossing, our balloon averaged 127 knots from take off to landing and in the middle of the Pacific it sustained a speed of over 200 knots for several hours.

Jetstreams are charted by the Met Office using a variety of information such as weather satellites, aeroplane reports, met balloons and Doppler radar.

Distribution of weather information has been made a lot easier by the electronics revolution and it is today possible to receive in the capsule, detailed weather charts momentarily after they have been released by the Bracknell Weather Centre.

3. Logistical and Design Problems and their Solutions

The logistics and equipment will be tested in a European flight that will in itself be record breaking, putting emphasis on the ergonomics and efficiency of the capsule, together with mechanical and navigational performance.

The actual circumnavigation of the world will take place when every element has been tested and when the meteorological conditions promise the best chances of success, which at present suggests the period from November 1995 to February 1996.

As already mentioned, Per Lindstrand and Richard Branson used hot air balloons for their record breaking flights in 1987 and 1991, but the problems inherent in building a hot air balloon capable of remaining aloft for such a journey made it necessary to consider the alternative of a helium filled balloon to achieve global success.

Unfortunately however, night time cooling of the gas would cause such a balloon to descend out of the jet stream, which is the only reliable method of maintaining the speed necessary to make the flight possible.

3.1 Solution

Amazingly the answer has been provided by Jean François de Rozier who evolved a combined hot air and hydrogen balloon after his first manned flight in a Mongolfier hot air balloon in 1783.

Two hundred years later his concept will be used with the much safer, inert gas, helium, which will be kept warm with propane burners during the hours of darkness to maintain altitude and speed by remaining in the jet stream.

The de Rozier principle, as it is known, will provide a controllable method of sustaining such a flight, whilst alleviating the need to carry the vast amount of ballast that a pure helium balloon would need or the impossible amount of fuel required for a giant hot air balloon.

By taking the earliest method of travel, that of harnessing the forces of nature and combining it with the latest materials and technology, the team hope to overcome the last great aviation challenge.

3.2 Envelope Manufacture

Creating a helium envelope 10 metres higher than Nelsons Column, capable of retaining 900,000 cu.ft. (25,000 cu.m.) of helium and keeping it there for 3 weeks is a major operation never attempted before. The longest a gas balloon has been airborne is less than 6 days so far and to triple the existing record will require the latest in fabric design and extreme attention to detail during manufacture.

The envelope uses a high tenacity polyamide fabric as its main structural element. This is coated with a polyurethane compound creating a gas barrier and a block for u.v. light. This fabric on its own constitutes a perfectly adequate material for a normal gas balloon but not for one which is to stay airborne for 3 weeks. To improve u.v. protection and to further reduce porosity, an outer layer of aluminised melinex film is added. This material has been specifically developed by ICI and is one of the most efficient gas barriers available anywhere in the world. This laminated material will reduce gas losses from the envelope by a factor of 10.

The joining of the envelope is carried out by thermal welding which avoids the use of any adhesives with the subsequent environmental hazards. The welding process must be carried out very carefully as a single hole about the size of a pin head could lose the balloon 3-4 days of endurance. Once the manufacture of the Global balloon envelope starts, the assembly hall takes the appearance of an operating theatre using clean room technology etc.

3.3 Balloon Fabric Testing

The fabric for the Global balloon has been developed specifically for this purpose. In an ideal world, we would build the balloon and flight test it. However, it is not possible to fly the balloon prior to the actual record attempt as it is very unlikely that it would survive a landing without damage to which would be added the high cost of the helium gas. It is therefore, imperative that the envelope material is tested in simulated conditions as near as possible to those of the actual flight, prior to the construction of the balloon.

We chose to use an INSTRON materials test system and we acknowledge this company's immense help and support in setting up the tensile testing machine and for advising us on a realistic test schedule. INSTRON is the world's largest supplier of materials testing equipment and we are pleased to say that this installation has made it possible for exhaustive analysis of balloon fabric properties.

One factor in the testing is that this balloon will fly at altitudes and temperatures that no other fabric balloon has ever flown in; it will also be exposed to the violence of the jetstream and accelerated to speeds in excess of 200 knots – needless to say this will put immense strain on the fabric. The INSTRON machine has an environmental chamber that can bring the fabric under test from -70^{o}C to +330^{o}C. The tensile jaws are driven by a computer and so it is therefore possible to simulate in our laboratory at Oswestry 2 weeks of flight in the jetstream at the correct temperatures and fabric stresses.

We believe that this is the first time such testing facilities have been available for balloon manufacture; they are producing invaluable data.

3.4 ICL's Involvement

Fujitsu ICL Trading was invited to provide the computers to be used by Lindstrand Balloons for the design and manufacture of the balloon, the capsule, burners and fuel system.

The computers comprise seven ErgoPro X450/90s together with 17 inch monitors. These are substantial configurations, each with a Pentium 90 processor, 32 Mb RAM, 1 Gb hard disc, sound card and quad CD ROM. In addition to the demanding design and analysis work where they are used in conjunction with AutoCAD software and NISA II stress analysis programs, the Fujitsu ICL computers are being used for the administration and management of the project.

The balloon designed to fly at 25,000 to 30,000 feet is complex; the stress will be considerably greater than at ground level where it can only be half inflated to allow for expansion as it rises through the atmosphere. Fujitsu ICL computers have played a key role in completing the design calculations in good time.

The cut out patterns for the balloon fabric, the design of the capsule, together with the burners and fuel systems to provide the hot air are all being completed on the Fujitsu ICL equipment. The capsule which has to contain the travellers for the duration of their flight must be as light as possible to minimise the energy required to hold it aloft.

4. Capsule Design

4.1 Pressurisation

In order to capture the jetstream the balloon must be able to sustain flight at 30,000 ft. Human life could not exist at this level without either breathing 100% oxygen or by creating a pressurised environment. While breathing 100% oxygen through a mask will give the required oxygenisation of the blood, other medical factors resulting from the lack of pressure in the body would bring great discomfort after 4-5 hours and hence this is not a practical proposition for a flight of long duration.

The only route feasible is to create a pressurised capsule and for this we had two choices. We could create a system like that of an airliner where air is pumped into the cabin continuously, in that case from the compressor section of the jet engine, and where the pressure inside the cabin is regulated by an outflow valve controlled by an aneroid control valve. The alternative is a closed system like that of a submarine or a spacecraft where the air inside the cabin is re-circulated and cleaned though chemical scrubbers and oxygen topped up through a liquid oxygen supply. The advantage with the latter is less weight and complexity but much lower air quality. The advantage of the former is a continuous supply of fresh air; it is however a highly complex system and also requires an air pump of some form.

We chose the former route just as we did with the Atlantic and Pacific crossings. We had to create our own air pump and we did this by running a propane powered piston engine on top of the capsule. This piston engine drives a large supercharger, which is a CompAir screw compressor, and this compressor (which consumes half the power of the engine) takes the air from the atmosphere and pumps it under high pressure into the cabin. This is a continuous process i.e. the engine and compressor are running all the time during the flight. The outflow valve regulates the internal pressure to a constant 8,000 ft level. As the air coming out of the compressor is + 165^{o}C, a series of intercoolers are brought into use in order to provide air at the correct temperature. The noise level immediately after the compressor is some 125 dB necessitating some clever silencing and technology to make the interior noise bearable.

As this system is crucial for flight at above 12,000 ft two identical engines/compressors are fitted to the capsule and both are run at half speed during the flight but in case of a failure of either unit, the remaining one can speed up and carry the entire load.

4.2 CAD - Computer Aided Design

This is an area where Fujitsu ICL computers have been able to play a key role. Initial design starts with 'Human Factors Engineering' the discipline involving the design of human related tasks, man-machine systems, ergonomics etc. Many elementary layouts are sketched based around the 'home' where the aeronauts will live and fly the balloon – the capsule. The pros and cons of each are discussed and, using information published from the NASA and Naval research programs, the designer configures a 'space envelope' based on the amount of working space required for flight control, space for stowage – equipment and food, area for rest and/or recreation, and most importantly, the amount of personal human space needed by each occupant.

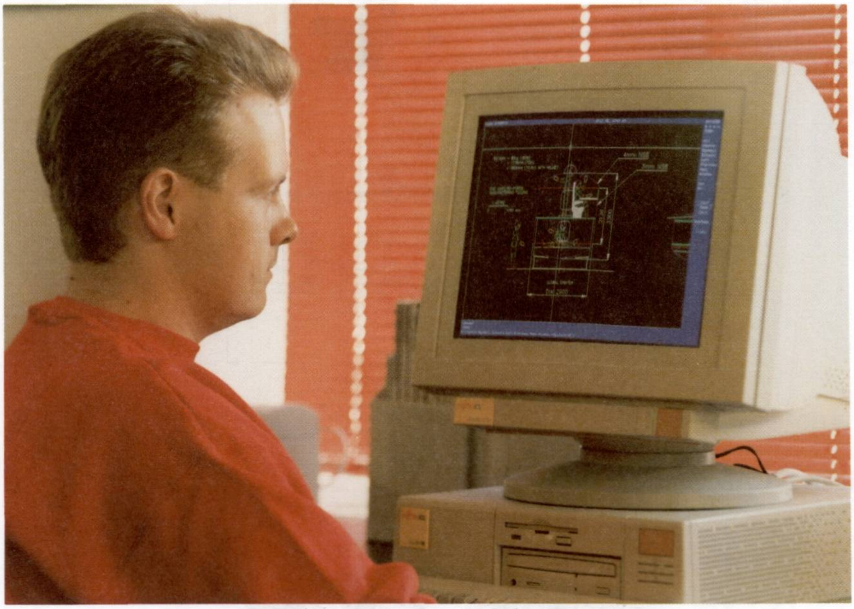


Figure 1 On-screen CAD design of the Virgin Global Challenger capsule

Design layouts are all initiated using a Computer Aided Design (CAD) program which allows many different configurations to be mixed and matched together rapidly. The sizes of any part of the structure or its area and volume are only a keypress away, saving many hours of laborious calculation. Three dimensional 'humans' are imported into the design scheme and moved into position to 'occupying' 3-D space. The structure can be quickly changed at will, parametrical stretching or shrinking sizes saving days of time compared with the traditional engineering 'tools' of pencil, paper and calculator.

Initial designs are downloaded from the CAD program for stress calculation to a Finite Element Analysis (FEA) program. The FEA results accurately display the stresses and deformation of the structure and any immediate problems are addressed before any further design is fruitlessly wasted. Confirmation of the structural integrity allows the detailed design to be stated at the earliest possible time.

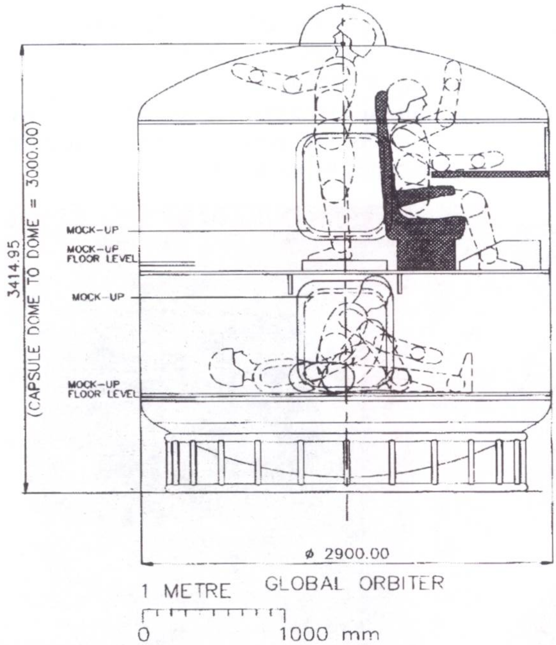


Figure2 CAD Capsule line drawing

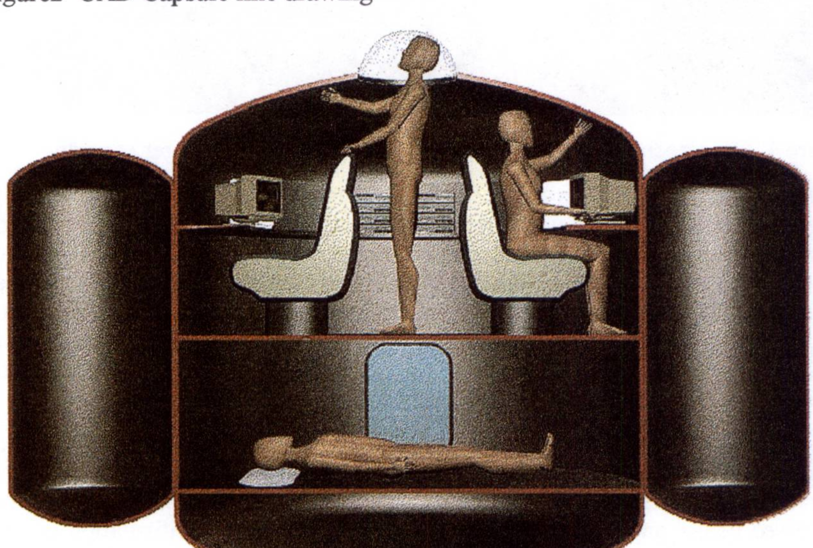


Figure 3 CAD Three dimensional 'human' design

With the capsule design finalised the weight of the structure and associated hardware fitments are calculated in 'real time' by the CAD program and the envelope can now be accurately sized to match the weight and lift requirements

Again the design file is downloaded from the CAD to the FEA system for initial analysis before finalising. The whole balloon system can be accurately modelled, 'built', 'inflated' and 'flown' at various altitude and climatic scenarios in a 'virtual reality' sense before even a piece of metal or material has been ordered or cut.

4.3 Stress Analysis

The role of the stress analyst within the Global project team is to ensure that the whole structure is safe to fly from a strength point of view. This not only includes the strength of the fabric envelope but also of the pressure capsule, propane fuel tanks, suspension load frame, seat attachments, etc. Furthermore, details of how the structure bends and twists are also to be calculated.

The complexity and quality of analysis for such a project is greatly reduced by employing sophisticated computer mathematical simulations by way of the well-known technique 'finite element analysis'. Finite element analysis grew up with the development of the computer and is today an indispensable tool for the major players in the aerospace industry.

Since much of what happens in the jetstream where the global balloon will be flying is unknown it is essential that we can predict the likely stresses and strains that the balloon will encounter so that we can prevent structural failure.

Several other areas of the flight spectrum need to be investigated with respect to the structural loads imposed on the balloon and capsule. For example the capsule will need to be assessed for pressurisation, launch loads, ascent, in-flight, descent and various landing scenarios on land or sea.

The structural analyst also needs to converse with the material specialists who provide breaking strength data. Where such advanced fabric materials are used in the construction of the global envelope, accurate test data is essential especially at extremes of cold to represent the conditions at altitude and extremes of heat to represent the propane burner heat effects on the fabric. The structural analyst needs to account for these factors for each particular load case.

Once this data has been assimilated it can be incorporated into the stress analysis model and run on the ErgoPro PCs. The output data is presented in clear colour graphics on 17 in monitors, making the task of structural assessment easier.

After consultation with the engineers on the team, decisions can be made from the model and incorporated into another run or runs until a sufficient level of optimisation and safety has been achieved and the design finalised for production.

5. Capsule Communications

An airborne aircraft normally communicates with ground stations with a VHF (Very High Frequency) transceiver. This frequency band is slightly higher than the broadcast radio VHF and while it provides for reasonable crisp and clear communication, it is limited to line of sight, as far as signal coverage goes. In a sense, if you can't see it you can't talk to it, which makes transmission over longer distances only possible when the aircraft is at high altitude. This doesn't normally cause a problem over land as there are plenty of transmitters and it is easy enough to install slave transmitters in remote areas. The limitations of VHF becomes acute during long over water flights as a typical jet aircraft cruising at 30,000 ft would be limited to a transmission range of approximately a 100 mile circle around the aircraft because of the line of sight limitations.

HF (High Frequency) radios can in theory transmit all the way around the world where the signal is propagated either along the surface of the earth called the 'groundwave' or by letting the signal bounce between the earth and the ionosphere. The 'groundwave' is limited to the power available on the transmitter and it is therefore practically limited to around 800 miles. The 'skywave' given sufficient transmitter power and the correct atmospheric conditions would allow this signal to go all the way around the world. In practical terms, it means it is possible for an aircraft in the middle of the Pacific to talk to either continental coast i.e. 3,000 miles away. The major disadvantage with HF is that the sound quality at best is poor and at worst almost incomprehensible.

In the last couple of years satellite communications have made large inroads and are likely to take over from HF progressively. An aircraft with this system will transmit directly to a satellite and the satellite will in turn transmit the signal back to a ground station. Of course, we are dealing with very high frequency digital technology and sound quality is extremely good. Obviously satellite time is very expensive but aircraft communications can usually be kept very short and more and more communications are now run by data transfer, further minimising satellite time.

The balloon will carry all three types of communication systems described above. Once the signal from the balloon has reached a ground station it will be patched (linked up) into a land line and from there to whomever the user wants to talk to.

6. Conclusions

The Virgin Global Challenger project has brought together many novel technical features which, it is hoped, will enable a flight duration of 500 hours to be achieved. A standard hot air balloon would be nearing the end of its useful life after this time, and would have had numerous careful inspections throughout its life to ensure its continuing safety.

The design of the envelope reflects the fact that this will be the first de Rosier Balloon to enter the jet stream. Simulated stressing techniques have shown how to optimise the structural design so as to accommodate the half-filled and fully inflated loadings.

The envelope material has been chosen as a result of considerable research into fabric and film composite structure covering the material performance in many diverse areas such as UV stability, helium leakage rates, infra red absorption/emittance, and structural strength. This work has also been extended to include developing differing joining techniques and full quality control of the resulting seams, since one failure in seam integrity, however small, will compromise the complete project.

The envelope control systems have also undergone re-examination. The gas release valve has been designed from scratch and tested for reliability in icy, sub-zero conditions. Smaller scale test envelopes have been built and flown in order to prove many of the above design aspects and to establish crucial performance data.

The capsule has undergone a similar process of design. Whilst most of the construction techniques are based upon established aviation technology, it is noteworthy that the projected duration of this single flight is believed to be considerably longer than that of any other pressurised aircraft. Consequently, the critical elements of the environmental control and flight systems have been duplicated to create a certain amount of safety through redundancy.

Extensive high altitude testing has been performed at Rolls Royce, Derby, to ensure that the pressurisation piston engines and hot air burners operate reliably at 30,000ft and that procedures can be established for re-starting stalled engines.

Answers for other, more obscure problems have had to be found, such as how to ensure reliable transmission and reception of satellite signals, when there is a 32metre diameter aluminised sphere directly above you. How to handle and inflate a 900,000cu. ft. envelope which weighs 1,500kg and then connect it to the capsule.

Potential answers have been found but the culmination of all our efforts will be the moment the Virgin Global Challenger lifts off and ascends to 30,000ft.

Acknowledgements

Martin Hutchins, Project Manager. Design engineer with design experience ranging from fighting vehicles to mini submarines and deep sea diving systems. A balloonist since 1970 with over 1000 hours logged. Project Manager with the successful Transpacific crossing in, 1991.

John Ackroyd, Aircraft designer with Saunders-Roe, Britten-Norman and Dornier. Vehicle designer with Porsche and Audi but foremost the designer of Thrust 2, the car that still holds the land speed record at 633mph. Capsule designer with the 1987 Atlantic crossing and the 1991 Pacific crossing.

Alex Ritchie, Gas turbine designer, ex chief engineer with Noel Penny Turbines, engine experience ranging from ships' diesels to drone engines. Mex was also part of the design team for the Atlantic and Pacific crossings, in charge of capsule pressurisation.

Andrew McCamley, Design engineer, ex Slingsby, BAe, Field, Airship Industries, designed airships for Per 1988-89 at Thunder & Colt. In charge of envelope design and manufacture.

Leon Eversfield, Structures engineer, senior lecturer at Kingston University, CAD and NISA expert, in charge of project stress analysis.

Baz Ferris, Design engineer, ex Rolls Royce, BAe, Slingsby, Airship Industries. Specialist in composite design.

Dr Janet Folkes, Materials technology. Degrees in geology, chemistry and metallurgy, laser cutting, in charge of materials testing.

7. Biographies of The Crew

Richard Branson

Richard Branson began his career at Stowe School at the age of 16 where he established a national magazine called Student. He then started a Student Advisory Centre to help young people and in 1970 he founded Virgin as a mail order record retailer shortly before opening his first record shop in Oxford Street, London.

He formed Virgin Atlantic Airways in 1984 and since 1985 he has been involved in a number of world record breaking attempts. In 1986 his boat 'Virgin Atlantic Challenger II, rekindled the spirit of the Blue Riband by crossing the Atlantic Ocean in the fastest time ever recorded. A year later he crossed the same ocean with the epic hot air balloon 'Virgin Atlantic Flyer' which was not only the first hot-air balloon to cross the Atlantic but was the largest ever flown at 2.3 million cubit feet capacity and reached speeds in excess of 130mph. In January 1991 he crossed the Pacific Ocean from Japan to Arctic Canada, the furthest distance of 6,700miles, again breaking all existing records with speeds of up to 245mph in a balloon of 2.6million cubic feet.

Per Lindstrand

Per Lindstrand was born in Sweden and moved to England in 1978 to start Colt Balloons Ltd; he has remained in England ever since, becoming firmly rooted in Oswestry.

His aeronautical career started with a short-service commission in the Swedish Air Force as an Engineering Officer. He later gained a masters degree in aeronautical engineering and worked for Saab Aircraft in Sweden and Lockheed in the USA. Whilst flying in the Air Force, he started ballooning as a hobby and his first balloon flight was a solo flight in his own home-built balloon. His interest in becoming a balloon manufacturer stemmed from his desire to extend the frontiers of balloon technology, as he observed there was a market opportunity to build and sell a properly engineered product in what was then more or less a cottage industry.

Colt Balloons' (and later Thunder & Colt's) trade mark was engineering excellence and that is what caused the Company to become internationally known. His interest in pushing through the boundaries of lighter-than-air technology also lead him to capture every absolute world record for hot air balloon flight. He is particularly known for the Atlantic crossing with Richard Branson in July 1987, bringing the distance record from 900 miles to 3075 miles. The following year he made a successful solo attempt on the world altitude record and flew his Stratoquest balloon in Laredo, Texas to 65,000 feet, bettering the old record by 10,000 feet. In January 1991 the Pacific Flyer balloon completed the longest flight in lighter-than-air history when Lindstrand and Branson flew 6761 miles from Japan to northern Canada, setting two new world records for distance and duration.

In December 1991 he founded Lindstrand Balloons Ltd. based in Oswestry and manufactures hot air balloons and airships, gas balloons and also maintains and repairs gas airships.

He flies not only balloons, but all other aircraft as well, having an airline transport pilot's licence for single and multiengined land and sea aeroplanes and helicopters and a commercial pilot's licence for autogyros and airships. He received the Royal Aero Club's Gold Medal from Prince Andrew twice, in 1989 and 1991, and the Royal Aeronautical Club's Britannia Trophy was presented to him by Lord Brabazon in 1988. He received of America's highest flying award, the Harmon Trophy, from Vice President Quail in the White House.

Author Biography

Simon Forse qualified in Aeronautical Engineering at Hatfield Polytechnic and first joined Thunder & Colt Ltd. in 1987 where he rose to the position of Head of Research and Development. Whilst his responsibilities covered every aspect of balloon equipment design, he specialised in burner and fuel system development. He designed the burners and fuel systems for the successful Stratoquest 65,000ft record breaking flight. For the 1991 Trans-Pacific flight, he again designed the burners and fuel system and also assisted in the balloon launch from Japan and co-ordinated the Japanese control centre.

In 1992 he joined Lindstrand Balloons Ltd. as Chief Engineer, where he is currently responsible for all aspects of engineering and certification within the company.

In 1994 he designed all the specialised equipment to assist Per Lindstrand and Judy Leden achieve the altitude record for dropping hang gliders at 40,000ft.

He began ballooning in 1978 and obtained his pilot's licence in 1982.

Further information

ICL Enterprise Technology, Multimedia Systems will shortly be displaying a number of pages on the World Wide Web, providing details of the project and progress both before and during the flight.

Design of the Format for EDI Messages Using Object-Oriented Techniques

Michael Chua

Independent Consultant, The European Commission, Luxembourg.

Abstract

With the advent of computers and communications technology, organisations are increasingly trading by electronic means, evolving the subject of *electronic data interchange* (EDI). Within the EDI development cycle, message design is an important phase as it dictates how effectively the business processes and objectives are met. The use of relational data models in message design does not provide the semantic richness at the technical and business levels that object-oriented models can easily achieve. This article describes the benefits of using object-oriented techniques in the process of EDI message design including some illustrative examples.

1. Introduction

Eurostat is the directorate in the European Commission responsible for the collection and dissemination of statistics in the European Community. To fulfil these responsibilities, Eurostat is developing an infrastructure for Electronic Data Interchange (EDI) to facilitate data exchange among the member states and Eurostat.

Nomenclatures are classifications of products and activities and form the framework of reference from which the collection of statistics is based. This is an important concern among the collectors of statistics, as reliable statistics need to be based on well conceived nomenclatures. Therefore, in order to harmonise these nomenclatures, there is the need for domain experts, nomenclaturists and statisticians to exchange opinions and data about products and activities and how they should be classified. EDI is the intended means to facilitate these exchanges.

2. Overview of EDI Development Cycle

EDI is a particularly wide-ranging application that is likely to influence most functions within an organisation and across every sector of industry, commerce and public administration. Organisations will be dramatically altered as they realise the power of EDI to exploit market opportunities, foster new business practices and stimulate the supply chain.

The UK National Computer Centre (NCC) has established a well-structured methodology (illustrated in figures 1 & 2 [NCC, 1992]) based on the practical experience of those who have already been down the EDI road. This methodology should enable all organisations to take full advantage of EDI. Managing EDI is not simply about developing one computer application within one department, it is about managing the growth of EDI throughout the organisation. It is about establishing a dialogue between trading partners such that each partner's strategy is developed and implemented to complement the other.

However, EDI is not a technology, it is a solution to business requirements. EDI merely applies other technologies like communications, computer hardware and software technologies to achieve the business objectives. Therefore, to achieve desirable results, it is necessary for EDI developers to possess a proper combination of technical and business knowledge. In an era full of technological marvels and advances, keeping up with the rapid changes in the basic technologies is crucial to understanding and realising the full potentials of EDI. Some recent examples include the proposed inclusion of multimedia formats for sounds and pictures in future EDI messages.

Unfortunately, the current trend of adopting Object-Oriented Design Methodologies in Software Engineering has not gained similar popularity in the EDI message design world. The object-oriented design concept applied to message design will result in better semantic richness at both the technical and business levels, as will be discussed later.

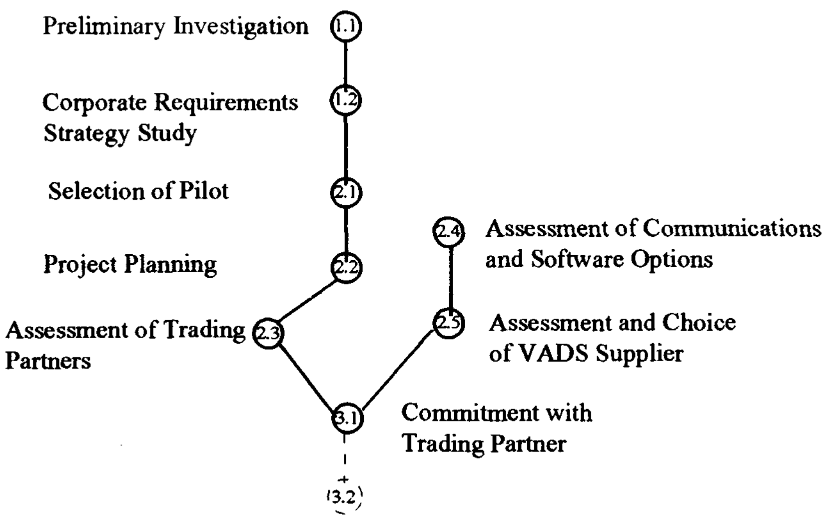


Figure 1 NCC Methodology Phases and Steps

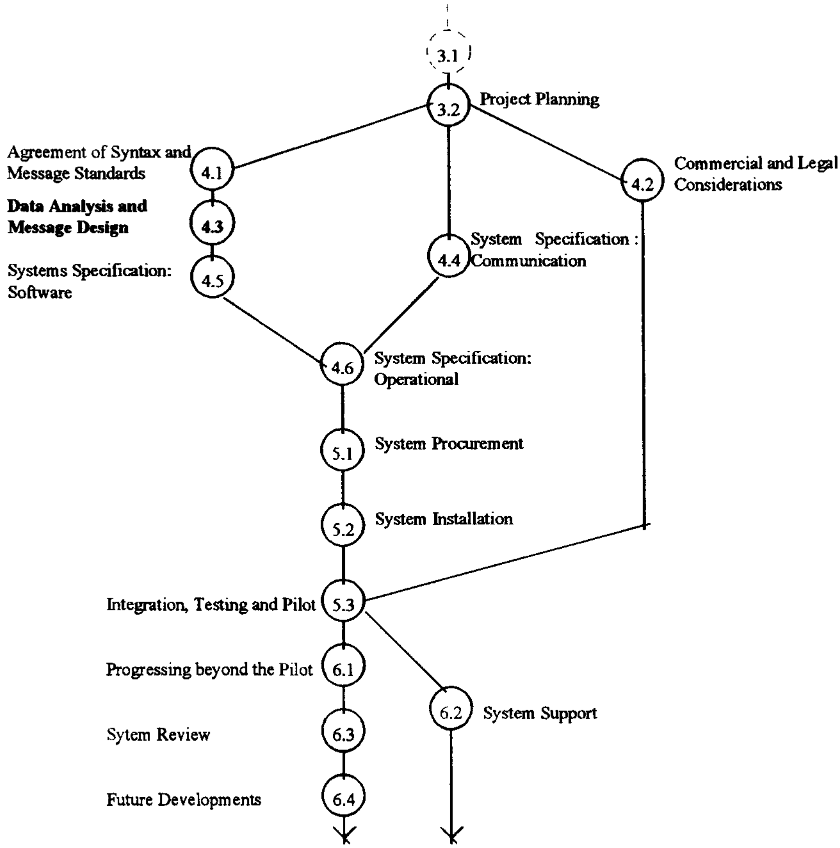


Figure 2 NCC Methodology Phases and Steps

3. Message Design by Relational Data Modelling and Intuition

In the early days of EDI, message designers relied mainly on their application expertise and intuition to construct their messages. During those times, due to the relatively limited technology, application requirements were more simple and modest and the lack of data modelling techniques did not cause a strain in message design. However, as in many other phenomena in life, the complexity of a problem becomes too great in practice once the scale and intensity increases beyond the grasp of intuition and application expertise.

The relational data model is an attempt to inject some structure into message design. Data are organised into entities and linked to one another by relationships. Though this is an improvement over mere intuition, it fails to capture accurately the semantic richness of the data at both the technical and the business levels. The logic of data transmission from node A to node B also dictates that data relevant to one another be closely coupled together. Hence, the transmission logic implies that messages are inherently object-oriented. Messages are not relational, as it is futile to transmit relationship indices which can be incompatible with indices in another node.

There are also other inherent weaknesses of the relational data model, particularly its effects after normalisation. Firstly, normalised relations rarely correspond to any object in the real world. The decomposition is dictated by computation or logic, rather than modelled on the nature and structure of the application. For relational modellers, it is instinctive to remove many-to-many relationships by introducing a new 'intermediary' entity which has no correspondence to the real world. Consider the relationship:

'Many cars may have many colours and many colours can be on many cars.'

To remove the many-to-many relationship, an artificial and 'intermediary' entity, called *colour mapping,* with no correspondence with the real world application is introduced, as shown in figure 3.

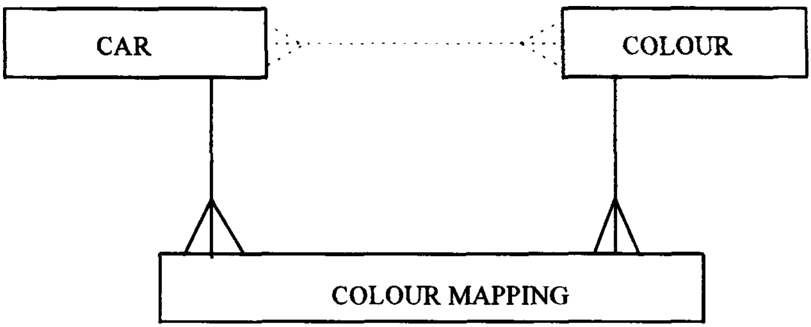


Figure 3 An example of an entity, *colour mapping,* that is introduced to remove many-to-many relationships.

So, in the relational data model, because of the lack of object identity, the tuples and tables may not correspond to a single object.

Consider for example designing an EDI message using the relational data model for the common vehicle accident insurance claim [Nelson, 1990], Figure 4 shows the entity-relationship (E-R) model and Figure 5 shows the message branching diagram[1].

In figure 4, each box in the E-R diagram represents an entity and each link represents a relationship. An entity is a 'thing', it can be abstract and may or may not correspond to real world objects. So, an object (in object-oriented modelling terms) is an entity, but the converse is not always true. Objects correspond closely to real world objects. In figure 4, *cost, damage* and *conviction* are not objects; they are entities.

[1] A message branching diagram is a component hierarchy showing the implementation details of a message.

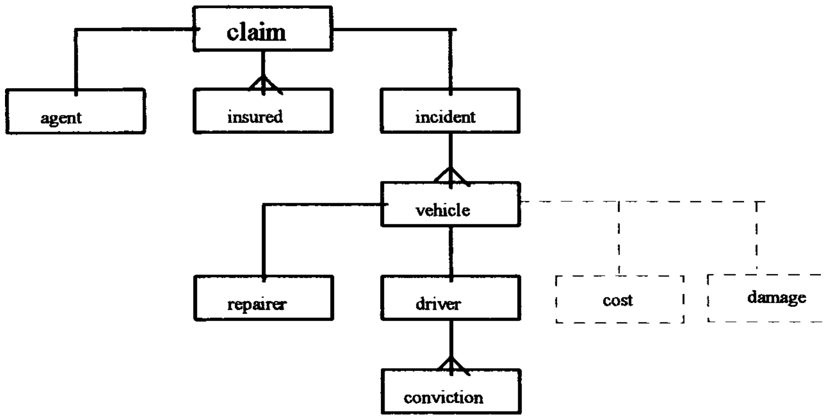


Figure 4 Entity-Relationship Diagram of a Motor Insurance Claim

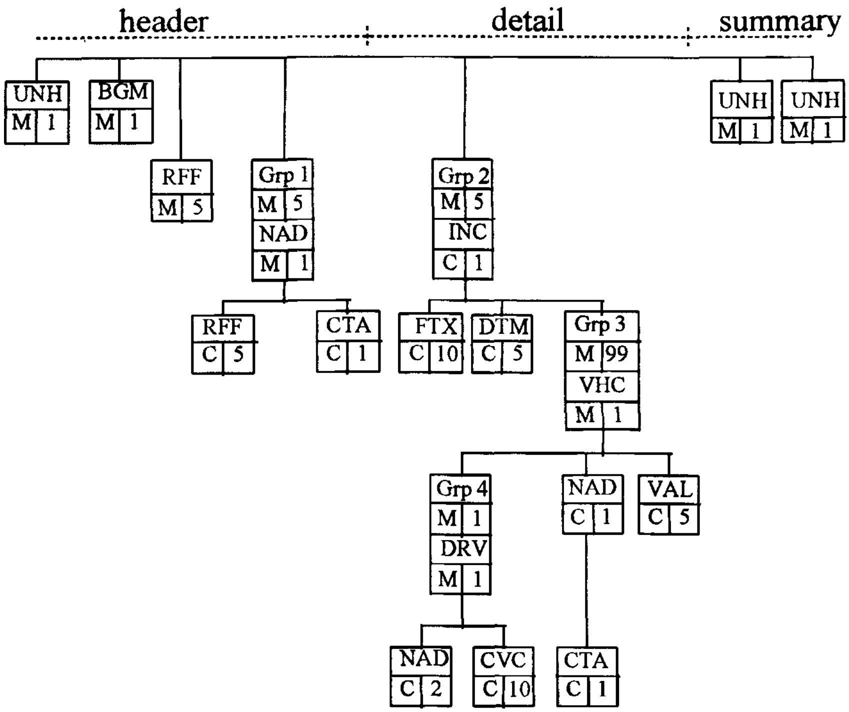
The author of this example assumed each vehicle involved in the incident to have one *cost* and one *damage*; and for this reason justifies making *cost* and *damage* attributes of vehicle rather than entities.

However, in a more realistic accident, there is usually more than one cost and damage for each vehicle, like car towing costs and car rental costs. And what happens if the accident involves damage to third party properties? According to the author, then we should have cost and damage as entities. But costing information for vehicle damage, car towing, car rentals and property damages are all very different. How are we going to reconcile these differences in entity *cost* and entity *damage*?

Evidently this E-R model is not extensible, because it does not set itself strict criteria for the determination of entities. The object-oriented model addresses these strict criteria. With unique object identities, it would also be possible to structure the *driver* and *vehicle* entities as independent and self documenting entities, higher up in the message hierarchy (figure 6).

Figure 5 shows the message branching diagram (using the UN EDIFACT syntax and standard segments), with a close structural resemblance to the E-R diagram. But with no strict criteria for the entities to correspond to real-world objects, it is likely that the entities are not self-contained or self-sufficient with its data attributes. Hence, the nestings of Group 3 under Group 2, and Group 4 under Group 3. This is structurally unwieldy. In an object-oriented model, each object would be relatively independent and would be less likely to need to be a sub-component of another higher component. The result would be a flatter message hierarchy more efficient for implementation and future extensions.[2]

[2] As mentioned in section 1.0, it is important that the message designer also appreciates the technical implications of his message, down to the level of programming in the wider EDI system.



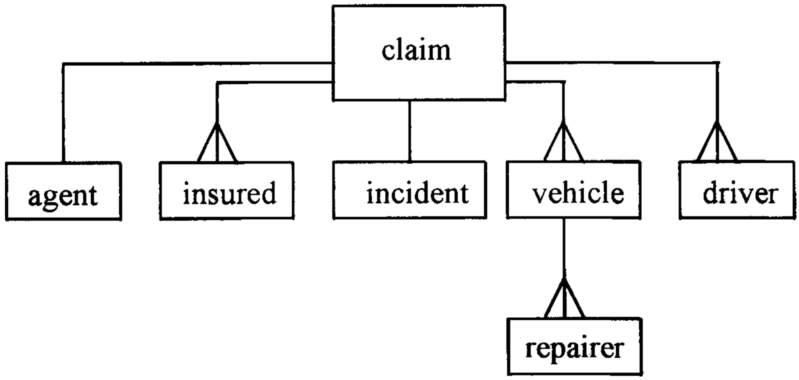
Note: The branching diagram has three parts, namely: the header which flags the beginning of the message and the identity of the sender; the detail part containing the application data; and the summary, flagging the end of message.

EDIFACT messages are designed with the principle of reusing existing standard segments. These segments have standardised abbreviated names, like NAD, for *name and address*. New segments can be invented, but they must be eventually approved by the EDIFACT board. Further details can be obtained from the EDIFACT documentation on standard directories.

In each segment, the small box in the bottom left corner specifies whether the segment is Conditional or Mandatory; and the bottom right corner specifies if the maximum number of possible occurrences of the segment.

Figure 5 Message Branching Diagram of Motor Insurance Claims

Unfortunately, it is still common practice to use relational data models to design EDI messages. Initially, thought processes are centred and based on relational data model concepts, often occupying a significant proportion of the message design phase. Only later when the satisfactory relational data model is achieved, do designers subconsciously think in object-oriented terms to convert the relational data model to a message branching diagram. It is during this unwieldy translation process that much of the even weak semantic contents of the relational model, represented by the links, are lost.



'Cost', 'Damage' and 'Conviction' are reverted as attributes.

Each vehicle can also have more than one repairer.

Figure 6 A suggested object-oriented data model

Therefore, for better efficiency and effectiveness, it is necessary for the data semantics to be systematically represented and documented. A better way of message design is to adopt object-oriented modelling.

4. Message Design by Object-Oriented Modelling

4.1 Object-Oriented Design Concept

Object-oriented modelling and design is a way of thinking about problems using models organised around real-world concepts.[3] [Rumbaugh et al, 1991] In the context of message design, we organise data as a collection of discrete objects that incorporate both data structure and behaviour. By contrast, in non-object-oriented design models data structure and behaviour are loosely connected, often dispersed in several data tables.

4.2 Object-Oriented Methods

It is not possible within the scope of this article to provide a treatise on the multitudes of object-oriented design methodologies practised in the industry today. To get a good introduction to many of the methodologies, see [Graham & Wesley, 1994].

Among the many methodologies in object-oriented design, this author has chosen to borrow the ideas from Semantic Object Modelling Approach (SOMA) [Graham, 1994], Class Responsibility and Collaboration (CRC) and Object Modelling Technique (OMT) [Rumbaugh et al, 1991]. These techniques will be discussed in the following sections.

---

[3] However, it is not always convenient to use purely real-world concepts, some abstract but discrete classes like trees and lists can also be useful

A detailed treatment of these three methodologies is not possible within the scope of this paper. The following sub-sections are extracts of the salient qualities of the methodologies which we intend to apply to message design.

4.3 Semantic Object Modelling Approach (SOMA)

SOMA [Graham, 1994] is chosen because it is a methods filter for other methods. It co-exists with other methods. But more importantly, it is chosen particularly for its use of Rules for object classes designed, an idea incorporated from the fields of Artificial Intelligence

SOMA is intended to be a semantically rich method for object-oriented analysis. SOMA can be used in conjunction with any other methods. In fact, SOMA started as an extension of the method and notation advocated by Codd and Yourdon [Codd 1970, 1983]. The distinct attractiveness of SOMA is that it involves clear steps to achieve full semantics for classes of objects. This semantic richness is achieved mainly by rules identification, a concept borrowed from Artificial Intelligence programming.

There are seven activities:

- identification of layers
- identification of classes
- identification of usage, classification and composition structures
- definition of data semantics and associations
- adding attributes to classes
- adding operations to classes
- adding the declarative semantics of the classes (rule-sets).

To provide the user with a concise understanding of SOMA in this limited space, the explanations that follow will be based on illustrative examples, particularly on Example 1.

Example 1: Nomenclature Problem (refer to Appendix for explanation of notations):

In a statistical context, nomenclatures are classifications of real-world objects which are used as a frame of reference in the collection, processing and publication of statistical data.

A Nomenclature is essentially identified by its name and date of release. The nomenclature item comprises primarily codes and labels and is also structured with the help of headings like section or chapters. There are also meta-data to help users to identify and relate to the nomenclature items, such as: owner of the nomenclature; its maintenance agencies; codes of related items existing in other nomenclatures; its validity period; keywords; and the data that enable nomenclature items from the same level, family or section to be identified. A nomenclature usually has one or more related nomenclature(s). There should be a transaction log to record access to the system.

4.3.1 Identification of layers

Object classes can be grouped into layers based on the interactions among themselves. The identification of these layers has more relevance in the dynamic behaviour of classes and is of lesser importance to message design. For a simplified methodology, this activity will not be emphasised here.

4.3.2 Identification of classes

Identifying classes of objects is an intuitive and subjective exercise. Apart from some basic rules, different people can have different opinions of what should or should not be a class.

As shown in figure 7, begin by listing candidate classes found in the written description of the problem. Do not be too selective; write down every class that comes to mind. Classes often correspond to nouns; verbs are candidates for behaviour or operations.

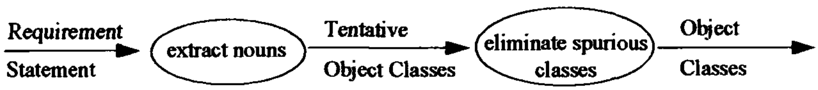


Figure 7 Process of identifying object classes

Do not worry about the composition structure of the classes, that will be dealt with later. First get specific classes right so that you do not subconsciously suppress detail in an attempt to fit a preconceived structure. Figure 8 lists the candidates for classes. From those candidates, we obtain a tentative list of classes illustrated in figure 10. Nomenclature classes may also be identified from knowledge of the problem domain, as in figure 9.

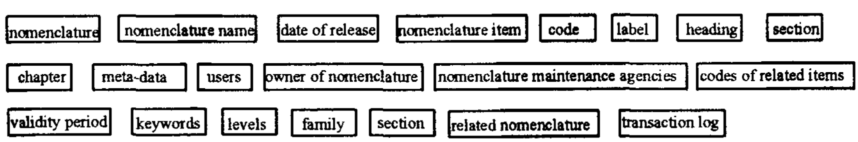


Figure 8 Nomenclature objects extracted from problem statement nouns

Figure 9 Nomenclature object identified from knowledge of problem domains

Irrelevant classes. If the class has little or nothing to do with the problem, it should be eliminated.

Attributes. Names that primarily describe individual classes should be restated as attributes.

Implementation constructs. Constructs extraneous to the real world should be eliminated from the analysis model. They may be needed later during design, but not now.

Redundant classes. If two classes express the same information, the most descriptive name should be kept.

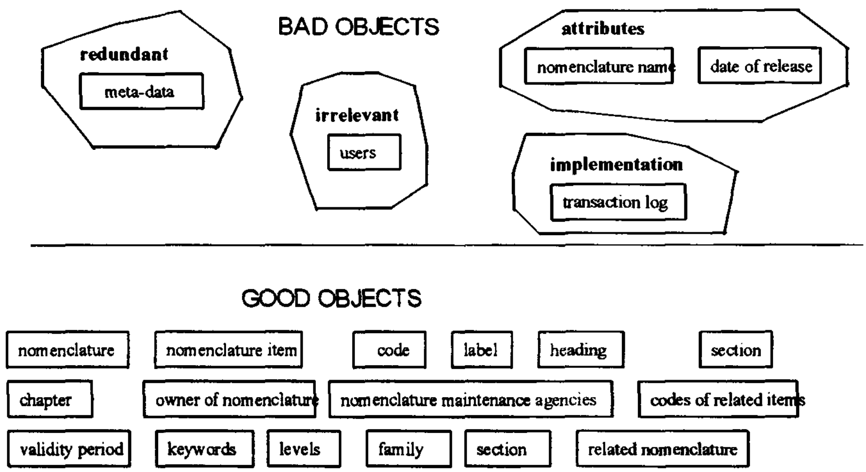


Figure 10 Eliminating unnecessary objects from the Nomenclature problem

Sometimes the nature of the project has a bearing on the class identification. For instance, in *owner of nomenclature*, there would be *address* which would contain an attribute *city*. In this case, it is sufficient that *city* exists as an attribute. However, in a Population Census Survey Project, City would exist as a class containing attributes like *name_of_city, number of inhabitants*, etc.

4.3.3 Identification of usage, classification and composition structures

4.3.3.1 Usage

In this phase, we attempt to identify the usage of each class. This will have a bearing on the definition of the class and hence its classification and composition structures. From the above list, it is observed that *nomenclature maintenance agencies* are used for informing organisations of who is responsible for maintaining the nomenclatures. The responsible agencies usually do not change frequently. Besides, including this class in an EDI message for nomenclature has little relevance as it does not contribute to the automatic updates of nomenclature contents. So, *nomenclature maintenance agencies* is useful information to keep, but should be excluded from the EDI message itself.

4.3.3.2 Classification

In Classification, the tentative classes are identified and grouped methodically. For instance, *section, chapters, level,* and *family* can be generalised as *groups*;

and *labels* and *headings* as *description*. However, as we generalise, the semantics of the class weakens. Conversely, as we specialise, the semantics of the class is enriched. However, the richer semantics for object classes keeps them more amenable to human thinking, but at the cost of a more rigid data structure.

Therefore, in classification, we have to determine the level of semantics we want to achieve. In the case of our Nomenclature message, we make a decision to generalise, as it has to be as generic as possible to accommodate any current or future nomenclature structures.

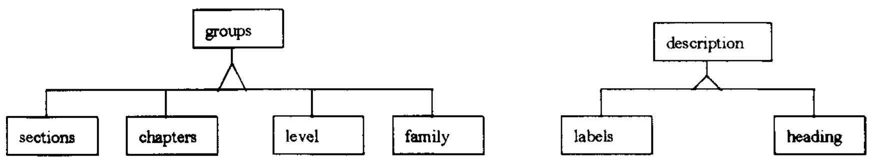


Figure 11a) Inheritance diagram[4] for entity *group* *11b)* entity *description*

4.3.3.3 Composition structure

Here, we organise the classes into a component hierarchy, as follows:

Note: let us call *codes of related items "links"*; and *related nomenclatures "relations"*.

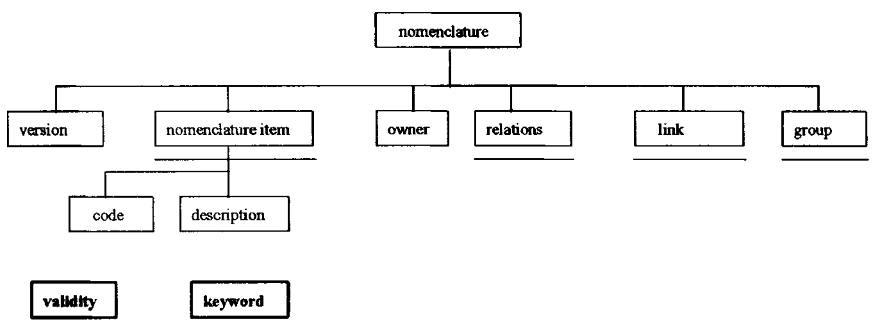


Figure 12 A partial component hierarchy for the message

Then, reiterating section 4.3.3.1, the usage of the classes, we discover that most of the classes require class *validity*. Class *validity* specifies the period of time that an object is valid. For instance, the code of a nomenclature item would have a validity period. Thus, the new composition of the message is illustrated in figure 13.

---

4 See appendix A Figure A2 for explanation of the sumbols used in the diagram

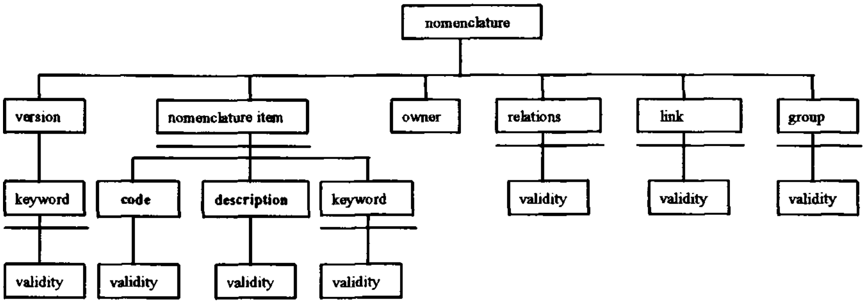


Figure 13 An initial component hierarchy for the message

Reiteration: Notice that classes *code* and *description* describe primarily the class *nomenclature* and according to 4.3.3.2, they should be considered as attributes of class *nomenclature item*. Also notice that *code* and *description* have a one-to-one relationship and hence can form a class together. Let us call this class *property*. The final result is the message component hierachy shown in figure 14.

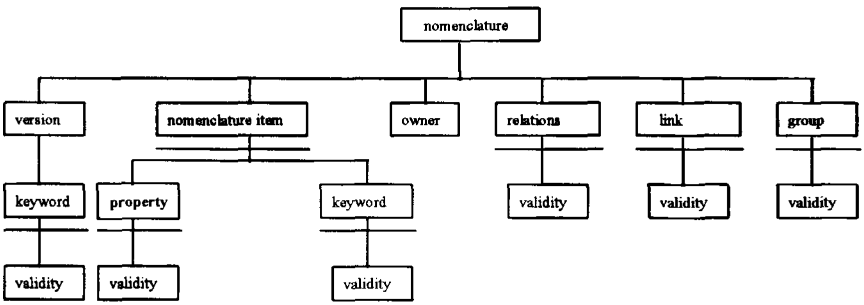


Figure 14 The final message component hierarchy

4.3.4 Definition of data semantics and associations

Class level connections is useful as it serves to enrich semantics. Connections of this sort may be thought of as associations: that is, structural relationships other than usage, classification and composition structures. In some cases these associations have properties and should be expanded into classes in their own right.

Incidentally, class *link* is one such case. It represents a nomenclature item relating to other nomenclature items. We have intuitively identified it as a class.

4.3.5 Adding attributes to classes

Here we take each class in relative isolation and fill it up with attributes. In 4.3.2 we have identified *name of nomenclature* and *date of release* as attributes. They can be included in the parent class *nomenclature* as attributes. In another class *validity*, we may identify attributes *start date of validity* and *end date of validity*.

4.3.6 Adding operations to classes

This aspect of the analysis is again more relevant to the dynamic aspects of the data. In the context of the message design and the limited space in this article, this section will not be explored.

4.3.7 Adding the declarative semantics of the classes (rule-sets)

This is perhaps the characteristic of SOMA that stands out against the other methodologies. The use of Rules for classes designed is an idea incorporated from the field of Artificial Intelligence; this involves identifying and specifying rules for each class. There can be global rules for a layer of classes for the whole system.

Table 1 illustrates the rules applicable for the class *validity*.

validity
Attributes
start date day/month/year
end date day/month/year
Rules:
1. end date >= start date
2. if month $\in$[4,7,9,11] then
day $\in$[1..30]
else if month = 2 and not leap year then
day $\in$ [1..28]
else
day $\in$ [1..31]
endif

Table 1 Specification of class *validity*

Adding the rule sets is the last activity in SOMA. However, analysis is an iterative process and it may be necessary to go back to the activities from 4.3.2 to 4.3.7 again, until a satisfactory solution is achieved. For instance, by going back to 4.3.2, within class *link* we might identify that there are two sub-component classes, namely: *source nomenclature item* and *target nomenclature item*, as illustrated in figure 15.

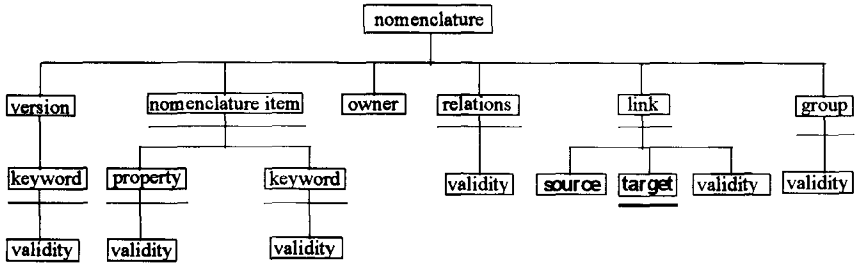


Figure 15 Source and Target Nomenclature items

4.3.8 Reusability and extensibility

Specifications that exhibit reusability and extensibility are all very well, but do not necessarily contain the meaning intended by the analyst or user.

To reuse a specification of a class, we should be able to read from it:

- what is the data structure
- what it does (operations)
- why it does it
- how it is related to other classes.

There are obvious advantages for having semantically rich classes. For one, if we are going to build executable and reversible specifications, it is crucial that information is not lost in coding. However, enriching the semantics of classes, in general, compromises reuse and makes the meaning of a class more and more specific.

4.4 Class Responsibility and Collaboration (CRC)

CRC is chosen for its simplicity and ease of use and its ability to co-exist with other methods.

CRC cards are useful for documenting object-oriented designs and also for teaching the basic concepts. This technique is often known as the responsibility-driven design (RDD). CRC can be made to co-exist with most of the other commonly practised methodologies. The unique advantage of CRC is that it uses nothing more expensive than a deck of paper index cards as a CASE tool. It is easy to delete the wrongly identified classes, attributes or operations – just rip the card and dump it in the bin!

The CRC card may look like the one in figure 16.

The usage of CRC cards may correspond directly with the seven steps of SOMA, particularly the initial phases of class identification. Throughout section 4.3, the reader can gain an intuitive understanding of its use by making a one-to-one substitution of the boxes in the diagrams with CRC cards. The links between boxes can be substituted by strings, or marked lines. The cards and the strings can be displayed on a white board with the use of magnets. This literal and open display of the classes and links enables group involvement in the design process.

Class:	
superclasses:	
subclasses:	
data:	
methods:	
used by:	

Figure 16 A template for a CRC card

4.5 Object Modelling Technique (OMT)

To date, OMT [Rumbaugh et al, 1991] is widely regarded as the most complete object-oriented analysis method so far published. It combines three views of modelling, namely: Object Modelling; Dynamic Modelling and Functional Modelling.

4.5.1 Object modelling

The object model describes the structure of classes of objects in a system – their identity, their relationships to other classes, their attributes, and their operations.

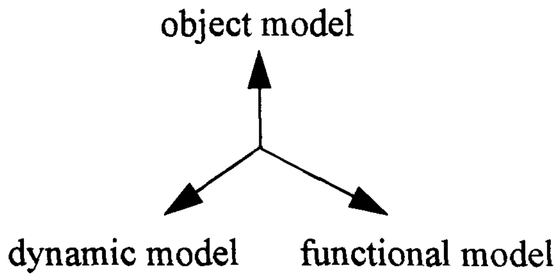
4.5.2 Dynamic modelling

The dynamic model describes those aspects of a system concerned with time and the sequencing of operations – events that mark changes, sequences of events, states that define the context for events, and the organisation of events and states.

4.5.3 Functional modelling

The functional model describes those aspects of a system concerned with transformations of values – functions, mappings, constraints, and functional dependencies.

In other words, OMT combines three different views of modelling. It is analogous to looking at a mountain from three different angles from which the mountain appears to be in three different shapes, but undoubtedly is still the same mountain.



Data-centred approaches to software engineering begin with the object model while process-oriented approaches start with the functional model. Some real-time approaches begin with the dynamic model. Data are more stable than functions and so the object model is the first approach towards modelling an information system in most cases.

With the three models, the message design is tested against the various requirements of the overall system, especially the subsequent step of systems specification. (refer to figure 1)

5. Eurostat's Nomenclature Database (SIMONE)

5.1 SIMONE Nomenclature System

In Eurostat and other statistical organisations, fields covered by nomenclatures include products, industrial activities, countries, currencies and occupations. The appropriate classifications are used at all stages of the statistics production process.

One of the objectives of [Simone, 1994] is to construct a generic system for construction, maintenance and consultation of nomenclatures; this means that the system must be able to deal with any form of nomenclature, official or unofficial. Consultation here means access by statistical systems as well as by people. The system must also be capable of interfacing with nomenclature dissemination tools.

5.2 SIMONE EDI Message

Subsequent to the analysis phase, two other classes of objects were identified for pragmatic reasons. They are *NAD* (name and address) and *nomenclature descriptor*. NAD is a standard data element from the EDIFACT directory; *nomenclature descriptor* is a suitable element to contain the identifiers and various administrative information.

The resultant message data structure is as illustrated in figure 17 and table 2.

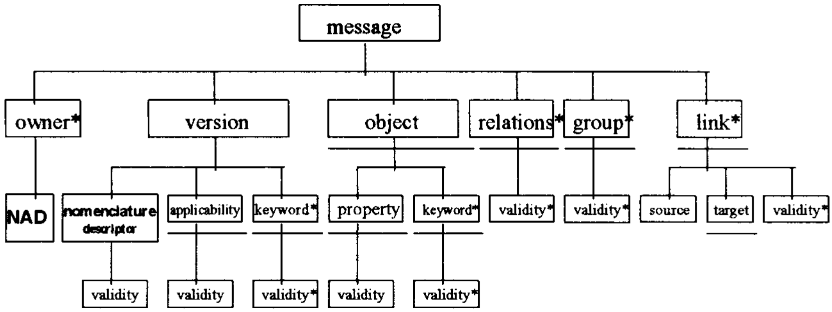


Figure 17 Component Hierarchy of the Message

Table 2 Summary of Object Classes in the Message

Data Element	**Description**
OWNER	Owner of the classification.
VERSION	Version of the classification.
NOMENCLATURE DESCRIPTOR	Describes the classification.
APPLICABILITY	Describes the class of economic activity.
KEYWORD	A list of keywords (for searching purposes).
OBJECT	A generic entity.
PROPERTY	Describes the property of OBJECT.
RELATIONS	Relationship among classifications.
GROUP	A group of OBJECTs.
LINK	Relationship among classification positions (items).
SOURCE	Source Classification Position.
TARGET	Target Classification Position.
VALIDITY	Date of Validity of parent entity.

5.3 Class Specifications

This section shows some sample class specifications of the Simone EDI message.

Table 3 Validity Classification

	VALIDITY
Meaning:	*Date of validity of parent entity.*
Attribute	
id	*identifier*
start date	*beginning of validity*
end date	*end of validity*
limit date	*lower limit of validity*
identifier:	*id*
Business Rules:	
end date > start date	
limit = max of (start date, end date)	*this is needed for data integrity checks.*

Table 4 Group Classification

	GROUP	
Meaning:	*A user-defined group of objects.*	
Relationship:	<Object > belongs to <Group>.	
Attributes:		
name		*eg. LEVEL2*
role	*role of the group*	*eg. level*
parent	*parent of the group*	*eg. LEVEL1*
VALIDITY	-	-
member code		
member name		
identifier:	*name*	
Business Rules:		
role $\in$	*[level, domain, field, family]*	
parent $\neq$	*Group.name*	

Table 5 Nomenclature Classification

	NOMENCLATURE DESCRIPTOR
Meaning:	*Describes the Classification.*
Attribute	
nomenclature name	*short name of nomenclature*
nomenclature title	*full name of nomenclature*
release	*date and a user defined text*
language	*specifies the language*
status	*specifies the status*
periodicity	*frequency of revision*
Identifier:	*nomenclature name*
Business Rules:	
nomenclature name ∈	*[NACE Rev N, ISIC Rev N, CPA, CPC, CN]*
nomenclature title ∈	*[Statistical Classification of Economic Activities in the European Community, International Standard Industrial Classification, Classification of Products by Activity, Central Product Classifications, Combined Nomenclature].*
where:	
NACE Rev N	*Statistical Classification of Economic Activities in the European Community.*
ISIC Rev N	*International Standard Industrial Classification.*
CPA	*Classification of Product by Activity*
CPC	*Central Product Classifications*
CN	*Combined Nomenclature.*
language ∈	*[DA, DE, EL, EN, ES, FR, IT, NL, PT]*
where:	
DA	*Danish*
DE	*German*
EL	*Greek*
EN	*English*
ES	*Spanish*
FR	*French*
IT	*Italian*
NL	*Dutch*
PT	*Portuguese*
status ∈	*[draft, operation, draft]*
periodicity ∈	*[daily, weekly, bi-weekly, monthly, bi-monthly, quarterly, semi-annual, annual, biennial, triennial, irregular]*

Table 6 Link Classification

	LINK
Meaning:	*Relationships amongst classification positions.*
Relationship:	*<Classification position> is related to <Classification position>*
Attributes:	
id	*identifier*
type	*type of link*
SOURCE	*the source object*
TARGET	*the object linked to source object*
VALIDITY	-
weighting	*the weightage of each link nomenclature position*
group	*the group the nomenclature position belongs to.*
identifier:	*id*
Business Rules:	
type ∈	*[hierarchical, correspondence, predecessor/successor]*
source ∈	*[Nomenclature_descriptor.nomenclature_name]*
source code ∈	*[valid Property.code]*

6. Conclusions

Message design is an important phase in the EDI development cycle as it dictates how effectively the business processes and objectives are met.

Intuition and application expertise were sufficient to design messages in the early days of EDI when requirements were modest and limited by technology. Relational data modelling is an attempt to inject some structure into message design, but is not coherent with the object-oriented characteristics of messages. In the relational data model, the semantic content of the model is mostly lost in the process of translating it into the message branching diagram. This translation process can be eliminated with the use of an object-oriented model.

Object-oriented modelling offers richer semantics to the data at both the technical and business levels. It enables the conception of the message in real-world terms, making it more amenable to human thinking. In SOMA, there are clear steps: to identify the classes of objects, their usage, classification and composition structures; define the data semantics and associations; add attributes and operations; and to declare the rules to each of the classes. In OMT, the message is taken a stage further to see how the object model would fit into the functional model and the state-transition model. This is a useful 'walk-through' before proceeding to the next phase of specifying the EDI software.

The concept of object-oriented analysis and design (OOA/D) has been explained mainly using the Simone EDI message as an example. Clearly, this is context-driven learning, and though not ideal, is a compact presentation within this limited space. A more formal presentation using abstract representations would clearly show that OOA/D can be applied to EDI message design in general.

References

CODD, E.F. "A Relational Model for Data for Large Shared Data Banks" , *Communications of the ACM* 13, No. 6, June 1970. Reprinted in *Communications of the ACM* 26, No. 1, January 1983.

GRAHAM, A. Semantic Object Modelling Approach (SOMA), 1994.

GRAHAM, I. and WESLEY, A. "Object Oriented Methods", 1994.

NCC. "The EDI Implementors' Handbook", UK National Computing Centre Ltd., 1992.

NELSON, C. "EDI The key to business success", Eirtrade, July 1990.

RUMBAUGH, J., BLAHA, M., PREMERLANI, W., EDDY, F. and LORENSEN, W. "Object Modelling and Design", Prentice-Hall International, 1991.

SIMONE: "Project Background and Status", Eurostat A1 Document, 1994.

Biography

Michael Chua is an independent consultant with the European Commission in Luxembourg and is also co-authoring a book on object-oriented analysis and design. His present work involves the investigation of the work processes to manage and maintain nomenclatures in the Commission, including the design and use of EDI messages. He has professional experience from several countries in Europe, the Far East and Australia working in the telecommunications, executive information system, manufacturing, engineering, process control and oil exploration environments.

He has a Postgraduate Diploma in Computer Science (1987) and a Bachelor of Engineering from the University of Newcastle, NSW, Australia (1985); and a Diploma in Civil Engineering from the Singapore Polytechnic (1980). Currently, he is pursuing a part-time MSc in Information Technology from the University of Keele, United Kingdom.

email: 100142.1370@compuserve.com

Appendix

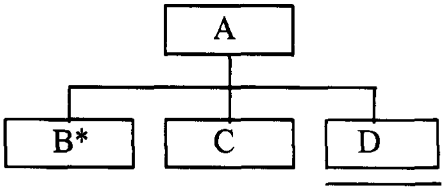


Figure A1 This means component A has one optional component B, one mandatory component C and one or more component D

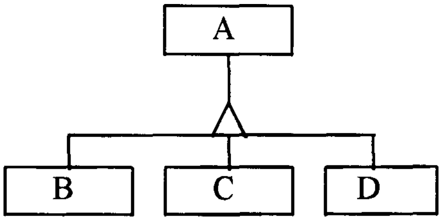


Figure A2 This means that class A is the parent class and classes B, C and D are derived from class A.

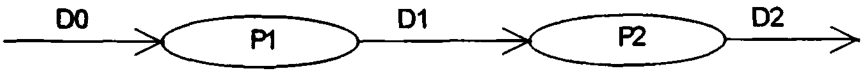


Figure A3 This a data flow diagram, where data set D0 goes into process P1 giving data set D1, which in turn goes into process

New Aspects of Research on Displays

Bo N. Schenkman

ICL, Sweden

Bo Persson

Institute of Optical Research, & Dept. of Architectural Lighting,

Royal Institute of Technology, Stockholm, Sweden.

Abstract

Current knowledge primarily about the ergonomic aspects of visual display units is briefly reviewed. It is expected that the use of displays will increase in a number of different situations. Emphasis is placed on seeing the display in a wider context, where environmental issues such as lighting are emphasised and considered in addition to task aspects and to individual aspects of the user. A proposal is outlined for an approach to research for the development of future displays based on perceptual ecology. Research should be more stressed on reading and writing processes for information design, as well as psychology.

1. Introduction

The basis for deciding if a display fulfils certain visual ergonomic criteria is the measurement and evaluation of a number of physical image quality parameters.

Which parameters these are, how they should be measured, and to some extent how they should be evaluated are described in different test procedures including international standards for displays, e.g. from ISO. Examples of such parameters are luminance contrast, flicker, jitter, raster modulation and glare.

Our involvement with developing international standards, in particular those for the Swedish National Board for Technical Accreditation, the SWEDAC/MPRII test procedures, relating to image quality of displays (these test procedures are discussed later in this article) have shown that:

1. The development and specifications of measurement methods relating to different image quality parameters are difficult tasks. One will find, for

example, that a specific parameter is measured in different ways depending on which international test procedure is used.

2. Relating a physical measured parameter to a relevant visual parameter e.g. physically measured luminance uniformity of a display to perceived uniformity of visual brightness of a display, is also difficult. This is complicated by the fact that the visual effect of all image quality parameters is affected by varying lighting and to some extent by individual characteristics of the user, such as age. For instance, younger persons generally need less light than older persons.

3. Stating which parameter, or which combination of parameters, is of major importance for a given task is itself complex. Such a decision depends on what is shown on the display. In other words, content of presentation is important.

The research effort on visual ergonomic aspects of displays must focus on solving the problems stated under 1, 2 and 3. This is a process which will take considerable time considering the fact that there are today no physical methods of measurement for visual brightness, the major concept concerning image quality of displays. A key reason for this is that a brightness measurement method must be based on one colour system, and today there are a number of different colour systems in use.

Research continually tries to improve or develop new parameters for image quality or test procedures, but the rapid technical development of displays, the rapid changes in how we use displays and the rapid changes in the environment where displays are used, is likely to set new demands on displays. It is therefore important to know and then to evaluate critically where visual ergonomic research on displays is heading and if there are new demands that should be taken into consideration.

Our article should be seen as a part of such a process, which points with some examples to rather new directions in the future design of displays.

For example, we want to avoid measuring and evaluating display problems that manufacturers are likely to solve in the near future. The reader will be presented with a context from which new and hopefully more visually and perceptually relevant test procedures for displays will emerge.

Information and entertainment will increasingly be presented on visual displays at work, in homes and for hobbies. Displays used at work are expected to become the dominating work tool for many different types of occupation. To be able to develop new displays and to understand the demands that will be made on future displays and on future display workplaces, new research strategies with new research methods are needed. The new methods must refer to the visual demands and needs for display users. Different users will demand different physical parameters to be fulfilled for a satisfactory image on the screen. These needs will depend also on the task to be accomplished. In perception research, one often uses the term perceptual ecology to describe the perceptual information that is presented to the observer [Gibson, 1979]. The display screen and what is presented to the computer user should, more than at present, be included in this visual world. From this new knowledge it will be possible to formulate improved and also new demands for future display work places.

2. Displays and their New Prerequisites

Information is presented increasingly by means of displays connected to a computer or network of computers. Even applications that used to need special measuring instruments can today be accomplished with a computer with appropriate hardware and software.

According to a number of indicators, the use of displays in working life is expected to increase. The number of different types of work where displays are the dominant work tool will increase. Displays will also increasingly be used outside of the traditional office environment e.g. for homework and for mobile applications. Added to this, there is an expected increased use of displays in spare time and for hobbies. We are today seeing a merging of computer technology, information technology and traditional TV technology.

There is a growing use of electronic games and of improved TV sets for ordinary TV watching. Many of the technologies in one field are transplanted into other fields, including computer displays. We may expect to see a continuation of this cross-fertilisation between technologies. This will also affect the expectations of users. Users will, for example, not accept that some screens are good for reading text, while others are better for viewing pictures or movies.

The possibility of being able to utilise and adapt to the information shown on displays will be important aspects of how satisfied users will be with their systems and displays. The adaptation to information will thus play a major part in determining the acceptance of computers and displays. By adaptation we mean the capacity of the system to change its parameter values to optimise its use, by taking account of the variability of the user and the environment. In view of this, the reading of text is an important aspect. [Wright, 1988] argues that reading and writing is a "pervasive thread" in working life. Written information in one form or another has an effect on almost all jobs. One has, for instance, to read manuals, machine instructions, warning signals, computer text and information bulletins. Wright uses the term "functional literacy" to refer to the information design and management needed in contemporary working life. She means that a person may be able to read, but they must be able also to use that ability to read and to look for information, especially on the new types of media today. Much reading at work is closely coupled with writing activities. This is not necessarily always the case with other kinds of reading, e.g. reading a novel for leisure.

The new possibilities to use information e.g. as a basis for decisions, will force us to an even greater extent to read text from displays. This type of information will and has already now taken such proportions that it is difficult and cumbersome to "print" the displayed text on paper. Future users of displays will probably demand that the information is presented in such a way that they can use the information without an unnecessary detour via printed paper. Special research has been aimed at finding out how best to design electronic documents and text see [Dillon, 1994]. He found that we often do not read from the first page to the last. There are many purposes for reading a text and the reading strategy depends on the purpose. He offers three criteria for classifying different types of text based on 'why' 'what' and 'how' characteristics of text as perceived by a reader. These are related directly to the design of electronic versions of the text. 'Why', 'what' and 'how' ask why the text is read, what type of information it contains and how it is read.

Apart from reading text, the display user will increasingly be confronted with new symbols and images that may be both static or moving. Film sequences and sound will be integrated with other presentation techniques. The current multimedia applications are waiting for their great breakthrough, in analogue with the breakthrough of the spreadsheet programmes in the beginning of the eighties. The route that multimedia will take will also have an effect on the design of screens. Will, for example, the new aspect ratios, i.e. the relationship of width to height, be 16:9 as for High Definition TV instead of the present common ratio for displays of 4:3?

On the basis of the perspective presented, we discuss some aspects of what should be done to acquire adequate knowledge for the development of future displays and also touch on how display work places should be refined in order to meet user demands for visual presentation.

3. The State of Knowledge Today

The technology that forms the basis for displays originates from a number of scientific areas. One of these areas is that of research on human perception and the visual system, an area that has been primarily the concern of psychologists and physiologists. An example of perceptual research affecting displays is the psychophysical models for flicker that have been proposed and used, [Eriksson & Bäckström, 1987]. Another example is work on a basic parameter for legibility, known as contrast, i.e. the quotient between foreground brightness and background brightness on the screen, see [Poynter, 1992] for a model of contrast sensitivity.

A second important source of knowledge is optics, from which important concepts and measurement procedures emanate. For example [Feng, 1994] developed and used concepts for describing image quality of lenses and photographs, when seeking to establish image quality parameters for visual displays. Feng adopted from the photographic branches of optics the concept of modulation transfer function area, which combines the physically measured sharpness of a display with a function for human visual sensitivity contrast. The modulation transfer function area is a Fourier transform of the luminance profile of an edge of a character weighted against the sensitivity to contrast. It had been shown that this function was useful as a quality measure for display images.

An area that is becoming increasingly important with reference to colour displays is colour itself. [van Laar & Flavell, 1990a, 1990b] give an overview on the problems of colour and also advise how to use colour in displays. An important factor which contributes to the confusion that plagues the study and use of colour in displays, arises from the large number of different colour systems used for describing colours. Among the colour systems used today are CIE and the Natural Colour System, (NCS) [Hunt, 1992]. ICL commonly uses the NCS notation for describing and defining surface colours of e.g. plastic materials, while the colours presented on the screen itself are defined by other colour metrics, in particular the chromaticity coordinates of CIE 1931. The colour system chosen should indicate how the colour will appear to an observer, and how different colours will look when viewed together at certain viewing conditions. One system may indicate that two colour samples are similar, while another might say that they will appear to look different from each other.

The large number of different colour systems that exist in the colour literature, puzzle not only a casual reader, but also make it difficult for specialists who are trying to establish adequate methods to measure colour characteristics of displays. The ideal colour system for displays would be a colour system based on how we perceive colours and specifically designed for luminous colours which are the type of colours seen on cathode ray tubes, CRT. Since there exists no such system today, we have to compromise by using colour systems that originally were created for other purposes.

A third area of knowledge that forms a basis for display technology is that of typography and aesthetics. However, it has been shown that this knowledge must be used carefully, since a number of methods useful for printed paper are not directly applicable to displays [Wright, 1988]. The following are some examples. Colours on tables on printed paper may assist readers, but not so when presented on a screen. Presentation of large tables as one single table might be appropriate for paper, but on a screen it is better through a number of smaller tables. Thus, information on a screen table might need to be duplicated in several places to make it more usable. The concepts of "before" and "after" have physical meaning for a printed page, but for an electronic document, in particular if based on hypercard technology, this may no longer be obvious.

Knowledge of perception, optics and aesthetics has formed a basis for the present dominant display technology based on CRTs. A good and basic overview of this technology is given by [MacArthur, 1980]. In his article a number of physical parameters for cathode ray tubes are listed. Among these are the character shapes and sizes, contrast, how the scanning of the electron beam is performed, the distortion of the picture on the screen, its stability, reflections and anti-reflective surfaces and choice of phosphors. Among safety and health aspects, Mac Arthur discusses the implosion risks, ionising radiation and high voltage. The heat emitted by cathode ray tubes is also mentioned.

3.1 A Swedish Perspective

Technology offers theoretical scope for possible solutions for displays. From this a suitable subset of available solutions must be selected and adapted to the needs of the display user. Research aimed at finding the visual demands that must be put on displays, and display workplaces, has for the most part derived from *ergonomic* aspects of the visual presentation. A number of Swedish institutes concerned with such diverse subjects as optics, lighting, perceptual psychology and environment, have for nearly two decades and together with the Swedish computer industry, in recent years especially ICL, acknowledged the importance of superior visual ergonomics in the work environment [Blomkvist & Schenkman, 1988].

The Swedish national board for technical accreditation, SWEDAC, earlier called MPR, has published testing procedures for displays that have had a major impact on how displays are evaluated. These testing procedures originated from a government initiative during the 1980s and encompassed both measures of the electromagnetic fields and of visual ergonomics. There is today an authorised testing procedure for displays [MPR 1990: 8, 10]. This testing procedure has had an international influence on display development and is today used as a de facto standard, although not formally a standard. The testing procedure has had a major effect on reducing the level of the electromagnetic fields emanating from a display. It has also been possible to compare different displays with respect to a number of aspects of image quality such as contrast and luminance uniformity. A formal standard for the electromagnetic fields is now under development by the Swedish Electrotechnical Commission, (SEK). One should also mention here that Swedish trade unions, e.g. [TCO, 1995], a white collar union, have shown a clear interest in promoting and working for good visual ergonomics in the work environment.

3.2 International Standards and Directives

Increased interest in displays can also be noticed in the development of different standards, where the present knowledge about and demands on displays are formalised. International standards for displays and their ergonomics are complete or under development. Of special note is [ISO 9241, 1992], which will consist of 17 parts, dealing with various aspects of the display workplace such as hardware, software, lighting and keyboard. In UK a BSI standard has been developed [BS 7179, 1990]; the part dealing with image quality of displays is in many ways similar to the corresponding part of ISO 9241. Both standards outline requirements for character size and shape, character spacing, luminance contrast, flicker, jitter and image uniformity. One difference is that the British standard includes the modulation transfer function area, mentioned above, as a measure of sharpness.

For displays, the VDU directive from the European Community, [Council Directive 90/270/EEC, 1990], plays an important role; it states in general terms how a display workplace should be constituted. This directive is primarily aimed at the employer, but it will also affect manufacturers, since the employer will demand that the directive is met. A list of minimum requirements is given concerning the display screen but also of keyboard, chair and desk. It is required for instance that the image on the screen "should be stable, with no flickering".

3.3 Increased Knowledge

Despite research conducted until now, our knowledge is not sufficient to guarantee new high visual demands regarding visual comfort and visual performance. We believe a number of causes to be responsible for this situation some of which are listed below. Their effects have been studied only to a small extent:

- more intensive use of displays by each user
- work that previously was flexible and variable for the user has become static and fixated in front of the display
- increased use of larger displays
- more information presented on the display
- presentation of moving images and sound
- increased use of colour displays
- use of new display technologies, e.g. liquid crystal displays, LCD, electroluminescent, and plasma displays
- increased use of displays in mobile applications
- increased demand of image quality on screen comparable to paper quality.

Our current knowledge is not sufficient for the scenario that we have described and there is thus a need to increase our knowledge of the demands put on display workplaces. This is needed to ensure that relevant demands are put on the crucial visual aspects. These are primarily the legibility of text, sharpness of images, absence of glare, absence of perceivable flicker or jitter, uniform movement of moving images, linearity and uniformity of objects and acceptable colour appearances. An important condition for this knowledge is a further understanding of the involved perceptual and psychological processes.

4. Increased Demands on Displays

The visual display unit was earlier regarded only as a technical tool intended for technically proficient users. With the spread of displays to new categories of users and to new types of applications, the image quality of the screen will be compared to the image quality of other media of presentation.

A computer and a screen may be used for a multitude of purposes and tasks. Many of these have no relation at all to reading text. However, when the screen is used for reading of text, the image quality will more and more be compared with that for print. This will demand that the present control methods for image quality are further developed, e.g by establishing new visually relevant parameters for image quality.

4.1 "Colour Shimmer"

One example of new concepts suitable for visual display units is the current work on defining a parameter we have called "colour shimmer". This is an effect, which can be seen on colour displays and on colour TVs, but does not occur on paper. The visual impression of this effect is caused by the geometrical layout of the red, green and blue dot structure of colour CRT and LCD displays. The visual effect might be described as a vague moving kaleidoscopic colour effect of the screen, that is best noticed when one tries to focus the eyes on the dot structure of the active display.

This effect will vary when the visual focus is shifted between parts of the display. The effect is different for different colour displays. We are so used to this effect, that we tend to overlook that it might become important when we compare e.g. the whiteness of a paper with the whiteness of a display. The effect is clearly shown when one simultaneously focuses on the structure of a well lit white paper placed over a part of the active display and on the structure of the colour display at a viewing distance of e.g. 30 cm. To assess the visual effect here called colour shimmer one needs an accurate and relevant physical measurement procedure.

4.2 Print, Luminance Uniformity and Sharpness

In order to be able to compare displays with printed material, we today lack relevant methods for such investigations. Studies performed by [Persson et al, 1993] show that current methods for visual display testing are not adequate for comparing displays. For example, according to the ISO 9241 standard, the important brightness uniformity of a display is measured physically as luminance uniformity at a number of geometrically specified points on the display. But what if the luminance variations occur between these specific points?

Another example is the need to measure sharpness of colour displays. This is covered neither by ISO 9241 nor by the SWEDAC/MPRII-test procedures. A visual inspection of a number of CRT colour displays will reveal the importance of this parameter. Sharpness might for example vary considerably over one active display area. It has been shown that there is a relation between physically measured and perceived sharpness of colour displays.

4.3 A Practical Example

From being viewed as a "technical product", which only is put on a desk, the visual display unit is becoming the prominent work tool for a number of occupations. The work environment must be built around this new type of activity. In the inter-disciplinary research project "Adaptive displays" a cooperative project between, among others, ICL and the Institute for Optical Research, a first step was taken to study quality demands in such a total perspective. Specifically, the project investigated how the image quality of a screen should be related to the varying ambient lighting in a room, in order to achieve comfortable reading of text on the screen. The project had a broad technical perspective, whereby a number of different disciplines were included, viz. optics, perception, lighting and physiology. The common denominator was the screen itself. It was established that visual comfort was a crucial factor for achieving the aims of the project. Visual performance is distinct from visual comfort, since the former focuses on some aspect of behaviour, e.g. time for recognition of a character, while visual comfort focuses on a feeling, i.e. how comfortable is the presentation?

The product that was realised, Auto BRITE, was based on the research methods worked out. It controls luminance on the surface of the screen as a function of incident light coming in front and from behind the visual display unit. The algorithm that describes the relation between the lighting in the room and the luminance of the screen was based on a number of empirical studies conducted with test persons. In these, a number of parameters, e.g. room illuminance, were systematically varied. A principle schema of the AutoBRITE system is shown in Figure 1.

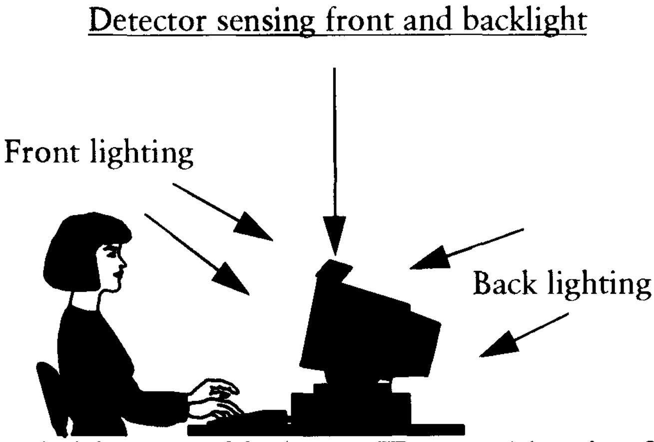


Figure 1 Principle schema of the AutoBRITE system: Adaptation of brightness and contrast in reference to varying front and backlighting

5. New Methods in Research and Development of Displays

New research on displays and their use can analyse more than what happens on the surface of the screen. One also has to relate this to the environment in which the display unit is situated. An important such environmental factor is, as has been noted, the lighting in the room. The need is established for research that seeks to understand both image quality and lighting, for example concerning the design of the active and passive parts of a display and their relation to different types of environment. To emphasise this, consider a beautiful or rare painting. The selection of frame and room for viewing the painting is critical to the overall visual experience of the viewed painting.

5.1 Influences from Room Lighting on Displays

The designer of room interiors where displays are to be used is guided by national lighting recommendations, which mostly state values of a few lighting parameters that are to be met. Examples of such demands are the required horizontal illuminance on a desk, e.g 500 lx, or that the maximum luminance in the surrounding areas should not exceed e.g. 200 cd/m

Added to these parameters are more generally stated proposals, e.g. that the luminance ratios (brightness ratios) of the display task, near surrounding to peripheral surrounding, should be 10:3:1 or 5:3:1. Such recommendations and proposals will at best provide basic lighting quality. One factor often overlooked is daylight which tends to give brightness or illuminance levels well above the stated levels. Another aspect is that some display workers prefer to work occasionally in a rather dark environment using only a desk lamp. The lighting recommendations do not cover such common lighting situations for display users.

It is of even greater importance to understand that a few physically measured lighting parameters will "describe" the actual lighting situation of a room only to a small extent. The room could be green or blue having a lot of shadows or no shadows at all. The window could be facing north or south allowing sunlight to enter the room. These factors, including sunshine, will have obvious effects on legibility, glare and the perceived colours of the display.

5.2 Influences from Typography

Users will demand a higher emphasis for aspects that previously were not attended to. Aesthetic aspects belong most probably to these new demands. Typographic knowledge and experience has had several centuries to find suitable and fitting forms for how text should be presented on paper, in news magazines, books, posters and other forms of printed material. Even today new forms are discovered or created. As an example of this continuous development, the Times character set, one of the most widespread today, was created in 1931 by the typographer Stanley Morrison [Hellmark, 1964] for the newspaper of the same name. Often, there are small deviations and alterations that may determine how aesthetically well-formed a particular character set appears. The corresponding knowledge for presentation on visual display units is making progress. For a comprehensive review, see [Rubinstein, 1988].

5.3 Suggested Research Approach

On the basis of the expected development of technology and the new demands from users of computers and displays, we believe that increased visual quality requirements will demand that a new research approach be established. This means research where perceptual and cognitive aspects on the use of display screens are studied in parallel with the establishment of relevant physical measuring methods. Basic for this research approach is:

- study of the user in his actual working situation, i.e. monitoring the user at work
- establishing relevant physical measures, e.g. measures that incorporate both physical and psychological factors
- measuring perceptual dimensions, e.g. visual comfort and visual performance
- relating the physical measures to perceptual dimensions, e.g. finding a suitable mapping of a perceptual dimension such as comfort to a technical specification such as contrast
- use of test persons with different requirements, characteristics and demands, e.g. experts vs non-experts, or young vs old persons.

The methods for this research approach are inter-disciplinary, in the sense that they use knowledge from physical science, behavioural science and liberal arts to gain technologically adequate knowledge. One major purpose is to attain knowledge about the user in behavioural terms, that can thereafter be translated into physical or technical concepts. These may later be implemented in a technical design that will form the basis for development of a product.

The approach we are advocating has been inspired partly by the perceptual ecological view proposed by the psychologist [Gibson, 1979]. Here one looks upon man and his behaviour as part of the ecological situation, in which he has developed. The display unit, taken together with the situation of the user, is according to our view, an adequate way for analysing future displays. An important unit for analysis will then be events of perception [Johansson, 1950]. These are the units of perception to which our visual system has been tuned by evolution. The visual system will, for example, see the common motion vectors in the optical flow that reaches the eye.

The ecological approach is an alternative theoretical view to the traditional classical approach for psychophysics. One way to express this classical view see [Zwislocki, 1965] may be described as follows:

$$\Psi = \Phi[f(s)] \qquad (1)$$

where

Ψ is the response function,
$f(s)$ is the stimuli for the sensory cells,
Φ is the response function in terms of $f(s)$.

We believe that one should express the concepts concerning man and display in terms of behaviour and situations, of which the user of visual display units is a part, rather than in terms related to physiological processes. The dependent variable Ψ, typically a subjective phenomenon, would thus be described as a function of non-reducible behaviour in certain situations or environments.

5.4 Future Studies

Among the display user issues that should be treated by the method proposed above, we will here mention some examples.

Understanding why reading text on a display is experienced as more difficult than reading on paper, is a basic problem for work with display units and for putting relevant visual demands on screens. It is primarily through research in the USA, that we have current knowledge about reading and writing and on how these processes are affected by computers [Wright, 1988]. Information design is emerging as a research domain in its own right. This new field receives input and generates output to related disciplines, e.g to cognitive psychology. Wright puts the emphasis on the process of writing, as distinct only from the final product, when she suggests an approach to the design of written information. It is important to understand the cognitive and perceptual processes invoked when people read and write – not only the objects they write on or read from.

The aesthetic aspects embrace physical, perceptual and cognitive dimensions. There is probably an invariant structure of aesthetics, but some of these aspects will also vary depending on time and culture. Some of the aesthetic experience may even reside in the newness of the phenomenon, i.e users may consider some design as beautiful, just because it is different and new, as compared with an older design, and not because of some inherent characteristics in the new design.

The causes of the difficulties of comprehending text on a display are to be found not only in concepts like character set, edge definition, character size, flicker or glare, but also in the passive parts of the display affecting how comfortable the user finds it to read from the screen. It is also an unknown issue how reading from paper, by reflected light, is different from reading from a screen, that uses emitted light.

For certain physical parameters, e.g luminance uniformity, a display user may accept variations of 10-20% over the surface of the screen. For other parameters, e.g. stray light from light surfaces/symbols on screens, the acceptability is only a few percent, as compared with the total brightness, the luminance dynamics of a depicted image. This variation sets high demands on the optical measurement technology.

6. Conclusions

The use of visual display units at work and in homes will increase. Future applications will be either modified or totally new. Research on visual ergonomics that should be performed, must satisfy new and increased demands that will be put on displays and work places for displays.

A new research approach with new methods and views has been outlined with the aim of gaining increased knowledge and from this a possibility to specify better and higher requirements for displays. Central to this research approach is a perceptual ecological perspective and knowledge about what affects visual comfort as well as performance, when reading text from a visual display unit. The approach is inter-disciplinary.

Even though the characterisation of future displays must be based partly on further improvements of current methods for testing image quality, the real challenge for future display developments will be in describing the "perceptual ecology" relevant for future display work. For a technically oriented reader such "soft concepts" might seem inadequate and not as readily understandable as image quality parameters. However, it has been found, for instance through the work on standardisation issues, that problems will persist if one focuses on technical aspects in only one field, e.g. on image quality, and forgets that there are other interacting fields that must be included.

Acknowledgement

For helpful comments, we thank Mr Gustaf Wottrich, ICL.

References

BLOMKVIST, A-C and SCHENKMAN, B. VDU ergonomics and office environment. Ericsson Review, 1, 1988.

BS 7179. Ergonomics of design and use of visual display terminals (VDTs) in offices. BSI, 1990.

COUNCIL DIRECTIVE of 29 May 1990 on the minimum safety and health requirements for work with display screen equipment. 90/270/EEC. Office Journal of the European Communities, L156/14, 1990.

DILLON, A. Designing usable electronic texts. Ergonomic aspects of human information usage. Taylor & Francis, 1994.

ERIKSSON, S. and BÄCKSTRÖM, L. Temporal and spatial stability in visual displays. In Knave, B and Wideback, P.-G. (Eds) Work with display units 86. Holland: Elsevier, 1987.

FENG, Y. Image quality criteria for photographic imaging systems and computer display screens. Licentiate thesis. Department of Physics, Royal Institute of Technology, ISSN 0280-316X, 1994.

GIBSON, J. The ecological approach to visual perception. Boston: Houghton Mifflin, 1979.

HELLMARK, C. Typografisk handbok. Stockholm, Ordfront & Ytterlids, 1964.

HUNT, R.W.G. Measuring colour. New York: Ellis Horwood, 1992.

ISO 9241. Ergonomic requirements for office work with visual display terminals (VDTs). ISO, 1992.

JOHANSSON, G. Configurations in event perception. Uppsala, Sweden: Almqvist and Wiksell, 1950.

MAC ARTHUR, I.D. Birds, Bs and CRTs, *ICL Tech J*, Vol. 2(2) pp.147-174, 1980.

PERSSON, B., ANDREN, B., HANSSON, L. and BREIDNE, M. Visual relevance of display image quality testing by photometric methods. TR 264, Institute of Optical Research, Sweden, 1993.

POYNTER, D. Contrast sensitivity and image recognition: applications to the design of visual displays. *Displays*, 13, 35-43, 1992.

RUBINSTEIN, R. Digital typography. New York. Addison-Wesley, 1988.

SWEDAC/MPR. Test Methods for Visual display units. 1990:8.

SWEDAC/MPR. User's handbook for evaluating visual display units. 1990:10

TCO'95 certification. Requirements for environmental labelling of personal computers. TCO, The Swedish Confederation of Professional Employees, 1995.

VAN LAAR, D. and FLAVELL, R. (a) How to use colours in displays- I. Physiology, physics and perception. *ICL Tech J.* Vol.7(1) pp.154-179, 1990.

VAN LAAR, D. and FLAVELL, R. (b) How to use colours in displays- II Coding, cognition and comprehension. *ICL Tech J.* Vol.7(2) pp.362-384, 1990.

WRIGHT, P. Functional literacy: Reading and writing at work. *Ergonomics,* 31, 265 - 290, 1988.

ZWISLOCKI, J. Analysis of some auditory characteristics. In Luce, R. D., Bush, R.R. and Galanter, E. (Eds.) Handbook of mathematical psychology, Vol III. New York, John Wiley, 1965.

Biographies

Bo N. Schenkman

Bo Schenkman gained a PhD in 1985 from Uppsala University in Sweden for research in experimental psychology. The same year he joined Ericsson Information Systems, that later became part of ICL. At ICL he has been working on human factors/ergonomics and related issues.

Bo Persson

Bo Persson gained an M.Sc in electronics in 1978 and a PhD in Architecture/ Lighting in 1990, both at The Royal Institute of Technology in Stockholm. He is currently working with research and commissions relating to image quality and lighting at the Institute of Optical Research and at the Royal Institute of Technology, respectively.

Book Review

"*Focus on IT in the City*", a report by members of the Worshipful Company of Information Technologists. June 1995 p.62 ISBN: 0 9526211

The Information Technologists are a new Livery Company founded in the 1980s that aims to apply the ancient traditions of Livery Companies in the City of London to the new art of IT. Edited by Alderman Sir Brian Jenkins, the then Master, Professor Richard Susskind, Mr Mike Warburg and Mr John Carrington, this report includes ten essays summarising the benefits a range of ten different forms of city business from banking to accounting have won from the application of IT. Each essay sets out succinctly and in non-technical terms how a particular sector has used IT, the chief competitive advantages obtained thereby, the threats and new opportunities IT offers and points at areas where further study would be valuable. Each was written by one or two authors with contributions from one to six others; altogether there are 13 authors and 35 contributors, all expert in their fields. The whole has been carefully edited and gives a balanced, up-to-date picture it would be difficult to match in so few words.

The Introduction by Professor Susskind calls for careful reading. The report was produced to give an overview of typical uses of IT by the City; it does not pretend to be comprehensive, but rather to be representative of advanced good practice. It is meant to allow broad comparisons with that adopted in overseas markets notably those of the USA. He asserts that the City is a world leading user in a number of sectors but that this lead is not permanently assured. He points to a need for lower costs, the establishment of standards and at the danger of excessive regulation. The papers agree on key technologies for enabling progress notably in electronic data communications, groupware, multimedia and the potential of the Internet for both dissemination of information and direct trading. Susskind sees the City as "a vast, sprawling collection of entities, relations and information flows, all charged with commercial significance".

The first essay is about *banking* starting with the familiar services provided by BACS (Bankers Automated Clearing Services) and CHAPS (Clearing House Automated Payments System) for cheques of ordinary and higher values respectively. The major preoccupations of the numerous domestic and foreign banks in London are with payments especially international payments and foreign exchange to which fast and totally secure IT is basic. Volumes are gigantic (CHAPS handles £100 billion daily) and competition is intense so that technical advances in security or speed of transfer will be rapidly taken up. The authors claim the City has people with the skills and experience needed to put such advances quickly into practice. Alongside exploiting technical improvements, bankers are working on quite new methods of cashless payment such as Mondex.

Next comes a discussion of *international trading in securities*, a field where London reckons to handle, through SEAQ International, 60% of all cross-border dealing in the world, again a very large-scale activity. The several commercial benefits of such electronic trading to participants in that market are briefly summarised. Though not all the operational and regulatory implications of this complex form of trading are explored in this essay, their importance seems sure to grow. Moreover, legal regulation is bound to require international agreement as recent well publicised events in Singapore and New York have shown.

Insurance is the next topic; London is typically a co-insurance market in which, as at Lloyd's, a number of underwriters may subscribe to a single policy thus calling for more data recording and processing. It too is an international market covering risks in all parts of the world. London has focused on world-wide risks to shipping and aircraft especially. Accounting is expected to become increasingly electronic particularly with the rapid growth of telephone sales of domestic and motor policies. Changes in working practices in insurance markets will promote wider use of IT, while failure to agree standards has been an obstacle to progress.

Fund management is yet another market in which international aspects are important particularly for institutional funds that have substantial foreign holdings to manage. Well known market reports (e.g. Reuters and Datastream) have long been widely used as have IT-based administrative systems. Financial modelling systems are used increasingly for investment decisions but the essay says London has fallen behind the USA in the use, for example, of such sophisticated techniques as neural nets and even chaos theory. A comparative study of American and British practice is suggested.

The challenge for *private banking* is to find ways of continuing to deliver the old-fashioned personal form of service that such banks like to give while making economic use of IT. Costs have fallen to the point where it may now be possible to do this. Private banks face radical alternative approaches, for instance by allowing customers direct access from home terminals to their accounts or to personal advice. It would appear that thinking on these issues is at an early stage.

Use of IT by the *legal profession* is obviously national rather than international because of wide differences in legal structures. English and Scottish lawyers, say, naturally seek to find outlets for their skills in those jurisdictions. Contracts between overseas parties from two or more countries may stipulate English law. Evolutionary rather than revolutionary changes are foreseen with wider use of general purpose IT systems to facilitate greater personal and collective efficiency in the office and in court. Despite the acknowledged conservatism of the profession the author sees definite scope for greater use of IT simply because administering the law is fundamentally an information handling process.

Accounting and audit have been among the oldest uses in the history of IT with roots as far back as the 1960s if not earlier. The essay, written from the view point of a major City accounting partnership, lists 10 distinct features of present day IT infrastructures in use amongst its clients. Accounting and auditing procedures sit naturally on top of routine data processing. One likely development is towards giving client staff greater access to the accounting and auditing functions – without offering any greater opportunities for malpractice. As with the law, international aspects are rather problematic.

Although provision of *specialised information services* on prices and markets is absolutely vital to the City, the next essay on Media and Information is less specifically restricted to the City than most of the earlier ones, simply because these services are taken by so many other organisations such as newspapers and broadcasters and, in the case of media, by almost any business. Since any item of news is liable to affect the stock market it can be hard to identify precisely the impact on the City of the use of IT by the media or to forecast trends. The author of the essay is sure, however, that the Internet with its "ability to carry information and financial services to business and domestic users" should be taken seriously.

The last essay on *Telecommunications and Communication Systems* is, as is the previous one, about the vital support those services give to the specialised sectors discussed in the first eight essays. Four layers of systems are distinguished, a) public dial-up systems as might be used by a private investor say, b) systems with access only over private lines and limited to subscribers , c) in-house or proprietary systems providing specialised analysis of certain forms of trading, and d) links between trading partners either in the City itself or around the world. Telecommunications has enabled dramatic changes to the way dealing is conducted; the so-called Big Bang led to the almost complete abandonment of dealing on the floor of the Stock Exchange. With the increased use of 'intelligent agents' to search for and process information to guide dealers, pressures grow for instant access to a wider range of databases not only those in-house but also of partners and service providers in the City and elsewhere. In international dealings seconds can count. London is a favoured hub for much international business data traffic but this lead is under constant threat. A key issue is the extent to which systems can be "open" to general access, while keeping the threat of fraud to an acceptably low level.

All essays compare the ranking of London favourably with that of other major financial centres; none sees it as unassailable since the rewards for successful support of financial business are so vast. Other factors than the purely technical, such as the skills of the individual operators and government regulation will determine which of a very few centres will come to dominate international financial activity. The authors of these essays obviously want London to lead and believe it can continue to do so granted proper support from IT.

They do not seem to be calling directly for the IT industry to define that support in business terms; rather they expect collaborative efforts led and inspired by the City itself to focus on what will be demanded by way of new forms of more intimate inter-business collaboration, much of it international. New international standards defining and regulating international trading will be framed. What makes this report so important to IT businesses such as ICL is that its authors plan to spell out, to the IT industry among other parties, what means of information handling and processing international financial businesses need to succeed.

JMMP

Subject Index
Volume 9

Volume 10

Author Index
Volume 9

Volume 10

Previous Issues

Vol.8 Iss.3 - May 1993

Vol.8 Iss.2 - November 1992

Vol.8 Iss.1 - May 1992

Vol.7 Iss.4 - November 1991

Vol.7 Iss.3 - May 1991

Vol.7 Iss.2 - November 1990

Vol.7 Iss.1 - May 1990

A Conversational Interface to a Constraint-Satisfaction System
SODA: The ICL interface for ODA document access
Human - Human co-operation and the design of co-operative mechanisms
Regulatory Requirements for Security - User Access Control
Standards for secure interfaces to distributed applications
How to Use Colour in Displays - 1. Physiology Physics & Perception

Vol.6 Iss.4 - November 1989

Time to Market in new product development
Time to Market in manufacturing
The VME High Security Option
Security aspects of the fundamental association model
An introduction to public key systems and digital signatures
Security classes and access rights in a distributed system
Building a marketeer's workbench: an expert system applied to the marketing planning process
The Knowledge Crunching Machine at ECRC: a joint R&D project of a high speed Prolog system
Aspects of protection on the Flagship machine: binding, context and environment
ICL Company Research and Development Part 3: The New Range and other developments

Vol.6 Iss.3 - May 1989

Tools, Methods and Theories: a personal view of progress towards Systems Engineering
Systems Integration
An architectural framework for systems
Twenty Years with Support Environments
An Introduction to the IPSE 2.5 Project
The case for CASE
The UK Inland Revenue operational systems
La solution ICL chez Carrefour a Orleans
A Formally-Specified In-Store System for the Retail Sector towards a Geographic Information System
...towards a Geographic Information System
Ingres Physical Design Adviser: a prototype system for advising on the physical design of an Ingres relational database
KANT - a Knowledge Analysis Tool
Pure Logic Language
The 'Design to Product' Alvey Demonstrator

Vol.6 Iss.2 - November 1988

Flexible Manufacturing at ICL's Ashton plant
Knowledge based systems in computer based manufacturing
Open systems architecture for CIM
MAES - An expert system applied to the planning of material supply in computer manufacturing
JIT and IT
Computer Aided Process Planning (CAPP): Experience at Dowty Fuel Systems
Use of integrated electronic mail within databases to control processes

Vol.4 Iss.3 - May 1985

Overview of the ICL Series 39 Level 30 system
VME nodal architecture: a model for the realisation of a distributed system concept
Processing node of the ICL Series 39 Level 30 system
Input/output controller and local area networks of the ICL Series 39 Level 30 system
The store of the ICL Series 39 Level 30 system
The high-speed peripheral controller for the Series 39 system
Development of 8000-gate CMOS gate arrays for the ICL Level 30 system
Development route for the C8K 8000-gate CMOS array
Design automation tools used in the development of the ICL Series 39 Level 30 system
Design and manufacture of the cabinet for the ICL Series 39 Level 30 system
Manufacturing the level 30 system I Mercury: an advanced production line
Manufacturing the Level 30 system II Merlin: an advanced printed circuit board manufacturing system
Manufacturing the Level 30 system III The test system

Vol.4 Iss.2 - November 1984

Modelling a multi-processor designed for telecommunication systems control
Tracking of LSI chips and printed circuit boards using the ICL Distributed Array Processor
Sorting on DAP
User functions for the generation and distribution of encipherment keys
Analysis of software failure data(1): adaptation of the Littlewood stochastic reliability growth model for coarse data
Towards a formal specification of the ICL Data Dictionary

Vol.4 Iss.1 - May 1984

The ICL University Research Council
The Atlas 10 computer
Towards better specifications
Solution of the global element equations on the ICL DAP
Quality model of system design and integration
Software cost models
Program history records: a system of software data collection and analysis

Vol.3 Iss.4 - November 1983

Expert system in heavy industry: an application of ICLX in a British Steel Corporation works
Dragon: the development of an expert sizing system
The logic language PROLOG-M in database technology and intelligent knowledge-based systems
QPROC: a natural language database enquiry system implemented in PROLOG
Modelling software support

Vol.3 Iss.3 - May 1983

IPA networking architecture
IPA data interchange and networking facilities

The IPA telecommunications function
IPA community management
MACROLAN: a high-performance network
Specification in CSP language of the ECMA-72 Class 4 transport protocol
Evolution of switched telecommunication networks
DAP in action

Vol.3 Iss.2 - November 1982

The advance of Information Technology
Computing for the needs of development in the smallholder sector
The PERQ workstation and the distributed computing environment
Some techniques for handling encipherment keys
The use of COBOL for scientific data processing
Recognition of hand-written characters using the DAP
Hardware design faults: a classification and some measurements

Vol.3 Iss.1 - May 1982

Software of the ICL System 25
Security in a large general-purpose operating system: ICL's approach in VME/2900
Systems evolution dynamics of VME/B
Software aspects of the Exeter Community Health Services Computer Project
Associative data management system
Evaluating manufacturing testing strategies

Vol.2 Iss.4 - November 1981

Architecture of the ICL System 25
Designing for the X25 telecommunications standard
Viewdata and the ICL Bulletin System
Development philosophy and fundamental processing concepts of the ICL Rapid Application Development System RADS
A moving-mesh plasma equilibrium problem on the ICL Distributed Array Processor

Vol.2 Iss.3 - May 1981

A dynamic database for econometric modelling
Personnel on CAFS: a case study
Giving the computer a voice
Data integrity and the implications for back-up
Applications of the ICL Distributed Array Processor to econometric computations
A high-level logic design system
Measures of programming complexity

Vol.2 Iss.2 - November 1980

The ICL Information Processing Architecture, IPA
VME/B: a model for the realisation of a total system concept
Birds, Bs and CRTs
Solution of elliptic partial differential equations on the ICL Distributed Array Processor
Data routing and transpositions in processor arrays
A Bayesian approach to test modelling

Vol.2 Iss.1 - May 1980

Security and privacy of data held in computers
CADES - software engineering in practice
ME29 Initial Program Load: an exercise in defensive programming
Project Little - an experimental ultra-reliable system
Flow of instructions through a pipelined processor
Towards an 'expert' diagnostic system
Using Open System Interconnection standards

Vol.1 Iss.3 - November 1979

Meteosat 1: Europe's first meteorological satellite
An analysis of checkpointing
Statistical and related systems
Structured programming techniques in interrupt-driven routines
The content addressable file store - CAFS
Computing in the humanities
The data dictionary system in analysis and design

Vol.1 Iss.2 - May 1979

Computers in support of agriculture in developing countries
Software and algorithms for the Distributed Array Processor
Hardware monitoring on the 2900 range
Network models of system performance
Advanced technology in printing: the laser printer
The new frontier: three essays on job control

Vol.1 Iss.1 - November 1978

The origins of the 2900 series
Sizing computer systems and workloads
Wind of Change
Standards for open-network operation
Distributed computing in business data processing
A general model for integrity control

To order back issues

Contact

Multimedia Solutions Help Desk

ICL, Westfields House, West Avenue, Kidsgrove, Stoke on Trent, Staffordshire, ST7 1TL, UK

Telephone +44 (0)181 565 7993 (ICL speedcall 7993)

Fax +44 (0)178 279 4870

Email:o.f.systemwise@kid0110.wins.icl.co.uk

Ingenuity

The ICL Technical Journal

Guidance for Authors

1. Content

Ingenuity, the ICL Technical Journal, has an international circulation. It publishes high standard papers that have some relevance to ICL's business. It is aimed at the general technical community and in particular at ICL's users and customers. It is intended for readers who have an interest in the information technology field in general but who may not be informed on the aspect covered by a particular paper. To be acceptable, papers on more specialised aspects of design or application must include some suitable introductory material or reference.

Ingenuity will usually not reprint papers already published but this does not necessarily exclude papers presented at conferences. It is not necessary for the material to be entirely new or original. *Papers will not reveal matter relating to unannounced products of any of the ICL Group companies.*

Letters to the Editor and book reviews may also be published.

2. Authors

Within the framework defined in paragraph 1, the Editor will be happy to consider a paper by any author or group of authors, whether or not an employee of a company in the ICL Group. All papers are judged on their merit, irrespective of origin.

3. Length

There is no fixed upper or lower limit, but a useful working range is 4000-6000 words; it may be difficult to accommodate a long paper in a particular issue. Authors should always keep brevity in mind but should not sacrifice necessary fullness of explanation to this.

4. Abstract

All papers should have an Abstract of not more than 200 words, suitable for the various abstracting journals to use without alteration.

5. Presentation

5.1 Printed (typed) copy

Two copies of the manuscript, typed 1½/2 line spacing on one side only of A4 paper, with right and left margins of at least 2.5cms, and the pages numbered in sequence, should be sent to the Editor. Particular care should be taken to ensure that mathematical symbols and expressions, and any special characters such as Greek letters, are clear. Any detailed mathematical treatment should be put in an Appendix so that only essential results need be referred to in the text.

5.2 Disk version

Authors are requested to submit a magnetic disk version of their copy in addition to the manuscript. All artwork and diagrams to be supplied in their original source format. The Editor will be glad to provide detailed advice on the format of the text on the disk.

5.3 Diagrams

Line diagrams will, if necessary, be redrawn and professionally lettered for publication, so it is essential that they are clear. Axes of graphs should be labelled with the relevant variables and, where this is desirable, marked off with their values. All diagrams should have a caption and be numbered for reference in the text and the text marked to show where each should be placed - e.g. "Figure 5 here". Authors should check that all diagrams are actually referred to in the text and that all diagrams referred to are supplied. Since diagrams are always separated from their text in the production process, these should be presented each on a separate sheet and, *most important*, each sheet must carry the author's name and the title of the paper. The diagram captions and numbers should be listed on a separate sheet which also should give the author's name and the title of the paper.

5.4 Tables

As with diagrams, these should all be given captions and reference numbers; adequate row and column headings should be given, also the relevant units for all the quantities tabulated.

5.5 References

Authors are asked to use the Author/Date system, in which the author(s) and the date of the publication are given in the text, and all the references are listed in alphabetical order of author at the end.

e.g. in the text: "...further details are given in [Henderson, 1986]"

with the corresponding entry in the reference list:

HENDERSON, P. Functional Programming Formal Specification and Rapid Prototyping. IEEE Trans. on Software Engineering SE*12, 2, 241-250, 1986.

Where there are more than two authors it is usual to give the text reference as "[X et al ...]".

Authors should check that all text references are listed; references to works not quoted in the text should be listed under a heading such as *Bibliography* or *Further reading*.

5.6 Style

A note is available from the Editor summarising the main points of style - punctuation, spelling, use of initials and acronyms etc. - preferred for Journal papers.

6. Referees

The Editor may refer papers to independent referees for comment. If the referee recommends revisions to the draft, the author will be asked to make those revisions. Referees are anonymous. Minor editorial corrections, as for example to conform to the **Ingenuity** general style for spelling or notation, will be made by the Editor.

7. Proofs, Offprints

Printed proofs are sent to authors for correction before publication. Authors receive either a hard copy master of their paper or a Word for Windows version in either V2 or V6 on magnetic media.

8. Copyright

Copyright of papers published in **Ingenuity** rests with ICL unless specifically agreed otherwise before publication. Publications may be reproduced with the Editor's permission, which will normally be granted, and with due acknowledgement.

Ingenuity November 1995