ICL TECHNICAL JOURNAL

Volume 8 Issue 2 November 1992

Published by
International Computers Limited
at
Oxford University Press

ICL TECHNICAL JOURNAL

The ICL Technical Journal is published twice a year by International Computers Limited at Oxford University Press.

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

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

1992 subscription rates: annual subscription £60 UK and Europe and $120 rest of world; single issues £36 UK and Europe and $72 rest of world. Orders with remittances should be sent to the Journals Subscriptions Department, Oxford University Press, Walton Street, Oxford OX2 6DP.

Printed by The Charlesworth Group, Huddersfield, UK, 0484 517077
ISSN 0142-1557

ICL TECHNICAL JOURNAL
Volume 8 Issue 2

Contents

Editorial Note

The Editor is greatly indebted to Dr. A.V. Flatman for considerable help in defining the topics and in finding the authors for the six papers on advanced networks in this issue.

Résumés

Jack Houldsworth
ICL Fellow, ICL Stevenage, Royaume-Uni
Reseaux ouverts — la cle du succes global

Cet article sert d'introduction aux autres textes consacrés aux réseaux dans ce numéro du ICL Technical Journal. Il commence par tracer l'historique de l'échange de données avec et entre ordinateurs, depuis ses débuts avec l'utilisation de méthodes hors ligne, puis de simples liaisons point à point par modem sur des lignes privées ou sur le réseau téléphonique public commuté.

L'article explique les raisons motivant le choix des techniques, telle que la commutation par paquets, qui se sont développées plus tard pour aboutir aux réseaux locaux et aux grands réseaux, et poursuit par une brève présentation des technologies actuelles et futures en matière de réseaux de communication à l'échelon mondial.

Cela conduit à un examen, d'une part, de la manière dont la standardisation des communications à un certain nombre de niveaux fonctionnels peut autoriser un interfonctionnement ouvert et, d'autre part, de l'architecture de communication bien structurée nécessaire pour permettre cet interfonctionnement ouvert.

L'article souligne l'importance du réexamen de l'organisation et des méthodes existantes des entreprises en tant que tout, présentant des stratégies techniques destinées à s'assurer que les séduisantes perspectives offertes par les formes les plus récentes de communication d'information de tout type, seront immédiatement exploitables dans la pratique.

Alan Flatman
OPENframework Division, ICL Kidgrove, Royaume-Uni
John Lawe
ICL Network Service, Stevenage, Royaume-Uni
Bernard Russell
Northern Telecom Europe, Londres N11, Royaume-Uni
Les infrastructures de réseaux d'entreprise dans les années quatre-vingt-dix

Les années quatre-vingt-dix verront und modification croissante des pratiques des entreprises, associée à l'application d'architectures informatiques réparties. Cet article examine bon nombre des nouvelles exigences sévères que ce changement imposera à l'infrastructure des réseaux locaux et des grands réseaux des entreprises. Les utilisateurs devront développer une stratégie réseau souple et ouverte, en sélectionnant la combinaison appropriée de technologies d'interconnexion pour répondre à l'évolution des besoins de leurs organisations.

L'article illustre et étudie des réseaux classiques, montrant la direction dans laquelle la forme de l'infrastructure réseau des entreprises devrait évoluer pour satisfaire les besoins futurs des utilisateurs.

John Meers
ICL Associated Services Division, Stevenage, Royaume-Uni
Réseaux large bande

Les réseaux de données jouent désormais un rôle essentiel dans la conduite de nombreuses opérations des entreprises. Les exigences de celles-ci en matière de réseau changent à mesure de l'évolution de leurs besoins. Alors que leurs exigences actuelles sont satisfaites par les réseaux à commutation par paquets et les réseaux locaux, les nouvelles applications utilisant des graphiques et d'autres techniques nécessitent une largeur de bande supérieure à celle autorisée par les réseaux existants.

Des techniques large bande offrant des débits de plus de 100 Mbit/s permettent aujourd'hui de satisfaire ces nouvelles applications. It existe actuellement un éventail de normes et de produits concurrents qui ont évolué à partir de techniques de réseaux existantes. La convergence vers un nouveau type de service reposant sur une allocation souple de la capacité commence toutefois à émerger.

A.V. Flatman
OPENframework Division, ICL, Kidgrove, Royaume-Uni
P.G. Weiser
ICL Logistics Division, ICl Stevenage, Royaume-Uni
L'évolution des réseaux sans fil

Les réseaux sans fil se révèlent une solution importante en tant que technologie de communication fondée sur des standards ouverts, tant localement au sein d'un bâtiment ou campus, que dans l'infrastructure de télécommunications à grande distance. Ils joueront de nombreux rôles dans les réseaux d'entreprise, qu'il s'agisse d'offrir une alternative au réseau câblé, de supporter des applications informatiques lorsque le câblage ne s'avère pas pratique, ou encore de compléter ce dernier pour répondre à des exigences particulières en matière de réseau.

Cet article étudie les développements des technologies, normes et règlements dans le domaine des réseaux sans fil, et envisage un large éventail d'applications potentielles avec les avantages qui y sont associés pour les entreprises. La terminologie utilisée est résumée dans l'annexe.

Mark Taylor
Memory & Technology Applications, ICL Manchester, Royaume-Uni
FDDI — le réseau grande vitesse des années quatre-vingt-dix

La norme de réseau local FDDI (Fibre Distributed Data Interface) qui émerge actuellement offre, en termes de performance, la capacité de pouvoir prendre en charge les utilisatuers de réseaux, au cours de la décennie '90. Cet article compare les performances de ce réseaux local de nouvelle génération aux technologies CSMA/CD et Token Ring actuelles et extrapole les tendances dans les dix années à venir. Il décrit de quelles manières le concept de réseau FDDI a évolué pour offrir une riche sélection de topologies et de technologies. Ces dernières permettent à l'utilisateur de composer un système adapté à ses besoins spécifiques tout en conservant tous les avantages de l'interconnexion de systèmes ouverts.

Edwin Turner
ICL Retail Systems — Bra 02
Technologie de communication pour l'environnement de la distribution

Bien que pour la plupart des gens, l'environnement de la distribution soit très familier, peu sont conscients de la technologie qui supporte les opérations de magasin au-delà des caisses et des terminaux point de vente (TPV).

Moins nombreux encore sont ceux qui apprécient la masse d'informations traitées et manipulées après la simple lecture du code à barres d'une boîte de conserve à la caisse du supermarché local.

Cet article étudie l'environnement dans lequel fonctionne l'équipement informatique en magasin et analyse quelques-uns des facteurs en limitant un déploiement plus important. Parmi ceux-ci, il est important de noter la difficulté d'interconnexion d'équipements informatiques dans cet environnement hostile. L'article explique ensuite comment la technologie de communication sans fil est prête à s'attaquer à cette tâche exigeante.

L.E. Crockford, A. Drahota
ICL
Riba — Environnement de support de traitement réparti

Aujourd'hui, les besoins informatiques des organisations de grande taille et de taille moyenne dépassent les capacités des systèms informatiques individuels et rendent nécessaire la répartition du traitement entre plusieurs systèmes. La plupart des organisations sont distribuées géographiquement et les utilisateurs d'ordinateurs au sein de celles-ci souhaitent des systèmes qui reflètent la nature répartie de l'organisation humaine. Il est rarement possible de séparer les données et le traitement sans aucune interdépendance entre systèmes. L'interfonctionnement est donc nécessaire entre applications tournant sur des ordinateurs centraux séparés.

La répartition du traitement introduit de nombreuses complexités pour les concepteurs, développeurs et administrateurs de systèmes. ICL développe l'environnement de support logiciel RIBA afin de réduire l'impact de ces complexités et, dans le même temps, de préserver l'investissement en développement effectué par ses utilisateurs. RIBA est fondé sur l'architecture ANSA et s'attaque à un large éventail de problèmes. Il est destiné à prendre en charge la répartition du traitement et des données dans des environnements hétérogènes d'ordinateurs centraux, de systèmes de gestion de bases de données, de langages de programmation et de systèmes de communication.

Etant donné la large portée de RIBA, cet article ne présente qu'un aperçu des approches adoptées, en particulier l'utilisation de mécanismes de transparence et d'outils de langage, et se concentre ensuite sur le thème de la transparence des données. Il s'agit là d'un développement RIBA qui cache au développeur la structure logique des informations implémentées dans diverses bases de données et qui permet de développer des applications en fonction de modèles d'information conceptuels. L'objectif de ce développement est de dissocier un programme d'application d'un type de base de données particulier et de permettre ainsi une migration future tout en préservant la portabilité de l'application.

ICL poursuit actuellement le développement de RIBA et plusieurs collaborations avec des groupements d'utilisateurs sont en cours. Elles démontreront l'opportunité et l'efficacité de la technologie RIBA dans des situations concrètes. Les composants de la plate-forme seront distribués en tant que produits ICL dans on proche avenir.

S. Blair Southerden, Manager, Investigative Systems
ICL Decision Support Product Centre, Bracknell, Royaume-Uni
Informatique: un outil pour les enquêtes et les renseignements judiciaires

Cet article décrit la base des techniques d'analyse de renseignements judiciaires dans les organisations chargées de l'application de la loi. La technologie permet désormais l'adoption et l'exploitation de ces techniques grâce au produit "Intelligence Analyst Workbench" d'ICL, qui met à la disposition des enquêteurs un système d'administration des connaissances.

Bruce Paterson
Normes en matiere d'agencement des claviers — les origines et la portee de la norme ISO/IEC 9995

Le standard ISO/IEC 9995, portant sur les configurations de clavier pour systèmes de texte et de bureau, est une nouvelle norme internationale qui devrait être publiée en 1992. Elle s'applique à l'agencement des claviers des produits informatiques, tels que les PC, stations de travail, terminaux et autres équipements. Elle remplace les normes antérieures pour les claviers de machines à écrire, machines à calculer et terminaux de télécommunication et de saisie de données. L'article étudie les points essentiels de la norme, évalue ses points forts et ses faiblesses, du point de vue des utilisateurs et des acheteurs potentiels de claviers.

En toile de fond à cette étude, l'article indique quelques faits majeurs dans l'évolution de l'agencement des claviers au cours des cent dernières années, et décrit les différentes tâches et rôles qui ont influencé la conception des claviers au cours de l'ère informatique. Il se termine en mentionnant brièvement quelques possibilités d'innovations bénéfiques dans la configuration des claviers.

A.W. Jenkins
ESS un système de disque à semiconducteurs pour les mainframes ICLSeries 39

Les systèmes de disque à semiconducteurs deviennent un élément essentiel dans la hiérarchie de stockage des systèmes informatiques importants. Ils sont indispensables pour résoudre le problème que constitue le décalage en termes d'accès engendré par la croissance des performances des processeurs qui ont déjà largement dépassé celles des systèmes de disque magnétique. Cet article décrit ESS (Extended System Storage), un système de "disque" à semiconducteurs développé pour les mainframes ICL Series 39. Il analyse les exigences prioritaires imposées au système et décrit les éléments de conception qui ont été inclus dans le produit pour y satisfaire.

Zusammenfassungen

Jack Houldsworth
ICL Fellow, ICL Stevenage, Großbritannien
Offene Netze — Der Schlüssel zum globalen Erfolg

Dies ist eine Einführung zu den folgenden Artikeln über Netze in dieser Ausgabe des ICL Technical Journals. Er umreißt zunächst die Entwicklungsgeschichte der Datenkommunikation mit und zwischen Computern seit ihren Anfängen, zunächst unter Verwendung rechnerunabhängiger Methoden und dann mittels einfachen Punkt-zu-Punkt-Verbindungen mit Modems über private Leitungen oder öffentliche Telefonwählleitungen.

Er erläutert die Gründe für die Auswahl von Techniken, wie die der Paketvermittlung, die später für den Aufbau lokaler Netzwerke und öffentlicher Fernbereichsmetze entwickelt wurde, und geht dann zu einer kurzen Einführung in die aktuelle und zukünftige Vernetzungstechnologie für weltweite Kommunikation über.

Dem folgt eine Untersuchung der Art und Weise, wie Kommunikationsstandardisierung auf mehreren funktionalen Ebenen eine offene Vernetzung ermöglicht und eine Analyse der gut-strukturierten Kommunikationsarchitektur, die für die Unterstützung dieser offenen Vernetzung notwendig ist.

Der Artikel betont, wie wertvoll es ist, bereits bestehende Betriebsorganisationen und-methoden ingesamt neu zu untersuchen, um zu gewährleisten, daß die interessanten Möglichkeiten, der heute vielfältigen Formen der Informationskommunikation auch in der Praxis ohne weiteres ausgeschöpft werden können.

Alan Flatman
OPENframework Division, ICL, Kidgrove, GB
John Lawe
ICL Network Service, Stevenage, GB
Bernard Russell
Northern Telecom Europe, London N11, GB
Infrastrukturen von Betriebsnetzen in den 90er Jahren

Die 90er Jahre werden eine immer größere Veränderung der Geschäftspraktiken bei der Anwendung verteilter IT-Architektur erleben. Der Artikel untersucht viele der neuen und anspruchsvollen Anforderungen, die diese Veränderung an die Infrastruktur der LAN- and WAN-Netze eines Unternehmens stellen wird. Die Benutzer müssen eine flexible und offene Netzstrategie entwickeln und die richtige Mischung aus Vernetzungstechnologien auswählen, um den sich neu herausbildenden Anforderungen Genüge zu leisten.

Anhand einer Beschreibung und Erläuterung typischer Netze wird gezeigt, welche Form die Infrastruktur eines Firmennetzes annehmen kann um die zukünftigen Anforderungen der Benutzer zu erfüllen.

John Meers
ICL Associated Services Division, Stevenage, GB
Breitband-Vernetzung

Die Datenkommunikation spielt heute eine Schlüsselrolle im Ablauf vieler Geschäftsabwicklungen. Mit den sich verändernden Anforderungen eines Unternehmens ändern sich auch die Ansprüche an das Netzwerk. Zur Zeit werden diese Anforderungen durch öffentliche paketvermittelende Netze und lokale Netzwerke erfüllt, doch neue Grafik Anwendungen und andere Übertragungstechniken benötigen eine größere Bandbreite, als diese Netze bieten können.

Breitbandtechniken mit einem Durchsatz von 100 Mbps und mehr können nun die Anforderungen dieser neuen Anwendungen erfüllen. Zur Zeit gibt es eine Reihe miteinander konkurrierender Produkte und Standards, die sich aus bereits bestehenden Vernetzungstechniken entwickelt haben, doch es beginnt sich eine Tendenz in Richtung eines neuen Services abzuzeichnen, der auf flexibler Kapazitätszuordnung basiert.

A.V. Flatman
OPENframework Division, Kidsgrove, GB
P.G. Weiser
ICL Logistics Division, ICL Stevenage, GB
Die Evolution drahtloser Netze

Drahtlose Netze kristallisieren sich zu einer bedeutenden Form der Kommunikationstechnologie mit offenen Standards heraus, und zwar sowohl auf lokaler Ebene innerhalb eines Gebäudes oder Betriebsgeländes, als auch in der Infrastruktur der WAN-Telekommunikation. Sie werden verschiedene Rollen innerhalb eines Firmennetzes spielen: als Alternative zu verkabelten, Netzen, als Unterstützungsmittel von IT-Anwendungen bei denen eine Verkabelung unpraktisch ist, oder aber als Ergänzung zur Verkabelung, um bestimmte Netzwerkanforderungen zu erfüllen.

Dieser Artikel analysiert die Entwicklungen drahtloser Netzwerkstandards und vorschriften und erwägt eine breite Palette potentieller Anwendungen mit ihren entsprechenden Geschäftsvorteilen. Die verwendete Terminologie ist im Anhang zusammengefaßt.

Mark Taylor
Memory & Technology Applications, ICL Manchester, GB
FDDI — Das Hochgeschwindigkeits-Netz für die 90er Jahre

FDDI (Fibre Distributed Data Interface) ist ein zunehmend wichtigerer LAN-Standard, der die Leistungsfähigkeit besitzt, Netzbenutzer bis weit in die 90er Jahre zu unterstützen. Der Artikel vergleicht die Leistung dieser neuen LAN-Generation mit den aktuellen CSMA/CD- und Token Ring-Technologien und projeziert die Entwicklung der Trends bis in das nächste Jahrzehnt hinein. Er beschreibt die verschiedenen Entwicklungsphasen, die das FDDI-Konzept durchlaufen hat, um eine breite Auswahl an Topologien und Technologien zu bieten. Diese ermöglichen es dem Benutzer, das Netzwerk seinen individuellen Anforderungen anzupassen und gleichzeitig alle Vorteile der offenen Systemverbindung aufrecht zu erhalten.

Edwin Turner
ICL Retail Systems — Bra 02
Kommunikationstechnologie für die Einzelhandelsumgebung

Die Einzelhandelsumgebung ist den meisten Menschen sehr vertraut, doch nur wenige sind sich der Technologie bewußt, die den Geschäftsbetrieb über Registrierkassen und elektronischen Kassensysteme (EPOS = Electronic Point of Sale) hinaus unterstützen. Noch wenigere haben eine Vorstellung davon, wieviele Informationen nach dem einfachen Abtasten einer Dose Bohnen an der Kasse von Sainsbury oder Tesco verarbeitet und für weitere Entscheidungen herangezogen werden.

Deiser Artikel analysiert die Umgebung, in der die IT-Geräte innerhalb des Ladengeschäfts arbeiten und diskutiert einige der Faktoren, die ihre weitere Entwicklung einschränken. Ein bedeutender Faktor ist die Schwierigkeit, die IT-Geräte in dieser schwierigen Umgebung miteinander zu verbinden. Sodann wird beschrieben, inwieweit die drahtlose Kommunikationstechnologie in der Lage ist, diese anspruchsvolle Aufgabe zu bewältigen.

L.E. Crockford, A. Drahota
ICL
RIBA — Unterstützungssumgebung für verteilte Verarbeitung

Heutzutage übersteigen die IT-Anforderungen großer und mittlerer Organisationen die Leistungsfähigkeiten individueller Computersysteme, und die Aufteilung der Verarbeitung auf mehrere Systeme ist eine Notwendigkeit geworden. Die meisten Organisationen sind geographisch verteilt, und ihre Mitarbeiter erwarten Systeme, die dieser verteilten Organisation entsprechen. Es ist kaum möglich, Daten und Verarbeitung ohne eine gegenseitige Abhängigkeit der Systeme zu verteilen. Aus diesem Grunde ist Arbeitsaustausch zwischen den, auf separaten Host-Computern laufenden, Anwendungen nötig.

Die Verteilung der Verarbeitung bringt viele Komplikationen für die Designer, Entwickler und Manager mit sich. ICL ist dabei, die RIBA-Umgebung für Software-Unterstützung zu entwickeln, um die Auswirkungen dieser Komplikationen zu reduzieren und gleichzeitig die Entwicklungsinvestitionen der Benutzer zu schützen. RIBA basiert auf der ANSA-Architektur und bewältigt eine große Anzahl von Problemen. Es zielt darauf ab, die Verteilung von Verarbeitung und Daten in heterogenen Umgebungen von Host-Computern, Datenbankverwaltungssystemen, Programmiersprachen und Kommunikationssystemen zu unterstützen.

Aufgrund des weiten Einsatzbereiches von RIBA bietet der Artikel nur einen Überblick über die verschiedenen Ansätze, insbesondere über die Verwendung von Mechanismen zur Datentransparenz und Programmierwerkzeugen. Dies ist eine RIBA-Entwicklung, welche in verschiedenen Datenbanken implementierte logische

Struktur der Informationen von dem Entwickler verbirgt und die Möglichkeit bietet, Anwendungen vor dem Hintergrund konzeptioneller Informationsmodelle zu entwickeln. Ziel dieser Entwicklung ist das Entkoppeln eines Anwendungsprogrammes von einem bestimmten Datenbanktyp, um so eine zukünftige Systemumstellung zu erleichtern und die Portierbarkeit der Anwendung zu bewahren.

ICL setzt die Weiterntwicklung von RIBA fort, und verschiedene Kollaborationen mit Anwenderorganisationen haben bereits begonnen. Diese werden die Relevanz und Wirksamkeit der RIBA-Technologie in tatsächlichen Geschäftssituationen zeigen. Komponenten der Plattform werden von ICL in nächster Zukunft als Produkte herausgebracht.

S. Blair Southerden, Manager, Investigative Systems
ICL, Decision Support Product Centre, Bracknell, GB
Informationstechnologie: Unterstützung von Untersuchungen und Informationsdiensten im Gesetzesvollzug

Dieser Artikel beschreibt den Hintergrund zu Methoden der strafrechtlichen Informationsanalyse innerhalb von Gesetzesvollzugsbehörden. Die Technologie ermöglicht die Übertragung und Ausschöpfung dieser Methoden mit Hilfe von ICLs "Intelligence Analyst Workbench" zur Lieferung eines Informationsverwaltungssystem für höhere Untersuchungsbeamte.

Bruce Paterson
Standards für Tastaturbelegungen — Entstehungsgeschichte und Umfang von ISO/IEC 9995

ISO/IEC 9995 Tastaturbelegungen für Text- und Bürosysteme ist ein neuer internationaler Standard, der 1992 veröffentlicht wird. Er gilt für die Tastaturbelegungen von IT-Produkten, einschließlich PCs, Workstations, Computerterminals und anderen Geräten. Er ersetzt frühere Standards für Tastaturen von Schreibmaschinen, Addiermaschinen, sowie Telekommunikations- und Dateneingabe-Terminals. Der Artikel bietet einen Überblick zu den wichtigsten Merkmalen des Standards und bewertet dessen Vor- und Nachteile in den Augen der Benutzer und möglichen Käufern von Tastaturen.

Als Hintergrund für den Artikel werden einige wichtigen Punkte in der Entwicklung von Tastaturbelegungen im Laufe der letzten hundert Jahre aufgezählt, zusammen mit verschiedenen besonderen Benutzeraufgaben und -rollen, die das Design von Tastaturen im Zeitalter des elektronischen Computers beeinflußt haben. Abschließend werden einige Möglichkeiten für nützliche Innovationen in der Tastaturbelegung umrissen.

A.W. Jenkins
Mai 1992
ESS — Ein "Solid-state"-Plattensystem für Großrechner der ICL Serie 39

"Solid-state"-Plattensysteme (RAM-Disks) sind zu einem zunehmend wichtigen Bestandteil der Speicherungshierarchie großer Computersysteme. Sie werden benötigt, um die Leistingsunterschiede auszugleichen, die dadurch entstanden, daß die heutige Arbeitsgeschwindigkeit der Prozessoren die Zugriffzeiten der magnetischen Plattensysteme bei weiten übertreffen. Dieser Artikel beschreibt die Erweiterte Systemspeicherung (ESS), ein für die ICL Serie 39 entwickeltes Produkt. Die wichtigsten Systemanforderungen werden analysiert und die Designmerkmale zur Erfüllung dieser Anforderungen beschrieben.

Open Networks – The Key to Global Success

Jack Houldsworth
ICL Fellow, ICL Stevenage, UK

Abstract

The paper introduces the others on networks that follow in this issue of the ICL Technical Journal. It first sketches the history of data communication with and between computers from its beginnings using off-line methods and then simple point-to-point links via modems over private lines or the public switched telephone network.

It explains the reasons underlying the selection of the techniques, such as packet switching, which developed later to build local and wide area networks and continues with a brief introduction to current and future networking technologies for world-wide communication.

This leads to an examination of the way standardisation of communication at a number of functional levels can allow open interworking and of the well-structured communication architecture needed to support that open interworking.

The paper highlights the value of re-examining existing corporate organisation and methods as a whole, outlining technical strategies to ensure that the exciting opportunities offered by newer forms of communication of information of every kind will be readily exploitable in practice.

1 Introduction

The telephone was demonstrated to Queen Victoria by Alexander Graham Bell at Osborne House, on the Isle of Wight, in 1875. There were readings of poetry from Windsor Castle and a piper played the 'retreat'. Her Majesty was most impressed but she could not have imagined that the new invention was destined to revolutionise both social and business cultures.

The first telephone exchange in Britain opened in 1876 with 8 subscribers and the rest is history.

Global networks now carry all types of business information within and beyond organisations. Voice and data co-exist quite freely and the services are rapidly being extended to embrace graphics, video and other kinds of business communications. Mixed media business applications will soon become as much a part of everyday life as the telephone.

This keynote article provides a brief introduction to current and future global network technologies.

It commences with a brief history of data networks, explaining how the current Local and Wide Area Network technologies were chosen, and shows how both current and future networking technologies fit together within a well structured open systems architecture.

The paper highlights the value of re-evaluating existing business cultures to make sure that the exciting opportunities which the new technologies can offer are fully exploited.

2 A Brief History of Networking

2.1 The Early Days

In the early days of computing, all the input and output devices were directly connected to the computer. Designers had enough problems with these locally connected devices without making the situation worse by trying to connect them from locations which were remote from the computer room.

In the late fifties, a few off-line transmission systems began to appear. They transferred information which had been captured on paper tape, or some other medium, to a similar peripheral in the computer room, thus avoiding the need to transport it physically. The final input to the computer was through one of the existing directly-connected peripherals.

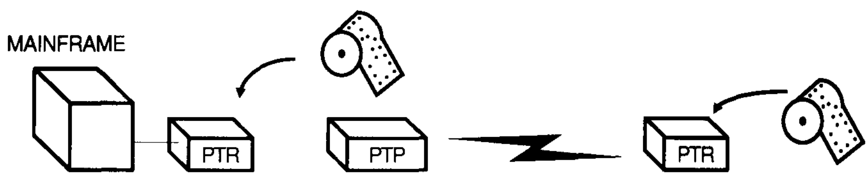


Fig. 1 Early "off-line" data transmission. PTR = paper tape reader; PTP = paper tape punch

Around this time, some directly-connected remote input/output devices began to appear. These were mainly electro-mechanical teletypewriters, running over leased telegraph lines, at speeds of around 10 characters per second. The author's first assignment on joining ICL in 1963 was to design the first teletype attachment to an ICL computer (the **original** ICT1500) – ICT was the name of ICL before the English Electric merger.

There were no data switching networks – other than the Telex network – and the telegraph authorities were very nervous about allowing computers to be attached to that. In any case, telegraph transmission speeds were obviously too low and attention was being turned towards using the telephone network.

2.2 The First Modems

Digital signals could not be transmitted through the voice network and adapters had to be designed to convert the data signals into voice frequency signals. These adapters became known as MODEMS (MOdulator-DEModulators).

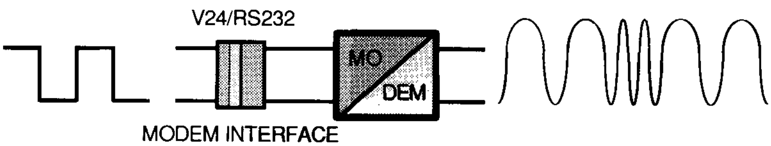


Fig. 2 Illustrating the frequency modulation technique used by a modem linked to another by a telephone circuit

The first modems appeared in the early sixties to support the off-line transmission systems, mentioned above. The first CCITT modem standard was ratified in 1968 but early versions of the same design had been in use for several years. It brought with it the first standard modem interface – CCITT V24 – often referred to by its USA reference RS232. This is still the most common interface around, particularly for connection to simple PCs.

The early modems were limited to transmission speeds of around 1200 bits/sec (150 characters per second), via either the switched telephone network or leased telephone lines, although 600 bits/sec was more common through the public network – contrast this with the 19.2 kbits/sec, and beyond, which is available today!

The adoption of modem standards meant that Global communication was assured. ICL demonstrated data transmission at 1200 bits/sec between the UK and Australia in 1968. All computer systems began to sprout connections to the outside world and networking began to blossom.

2.3 *The Introduction of On-line Systems*

The first on-line visual display systems to be connected via the telephone network were fast teletype emulators which copied the asynchronous teletype transmission system. Many simple screen and keyboard terminals still use this type of transmission and they are often referred to as teletype compatibles.

Computer communications systems soon began to use synchronous transmission with automatic error control and recovery. This was more efficient and reliable and soon became the preferred system for on-line computer networks. The first ICL machine to use this kind of transmission was the ICL1900 (connected via the 7010 Communications Controller) in 1967.

There were some protocol standards, based on the transmission control characters from the teletype code (now universally referred to as the ASCII 7 bit code). The standards offered a reasonable degree of compatibility but there were too many options to guarantee interworking between different manufacturers' systems. There was a lively business in gateways, which took care of the differences.

2.4 *Front End Communications Processors*

All the early on-line systems were star connected, with a mainframe as the focus. Any terminal-to-terminal information had to flow via the mainframe.

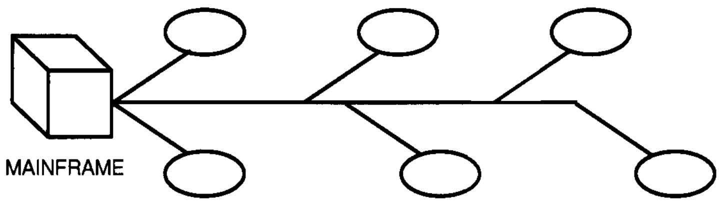


Fig. 3 A main frame connected to a number of terminals (shown as ovals)

Manufacturers began to introduce 'front-end' communications processors to handle very large numbers of transmission lines. They were soon adapted to switch the messages which were addressed to other sites, to keep them out of the mainframe altogether. Some users joined these front end processors together and used them as *nodes* in an embryonic data network.

As the numbers of lines and the overall volume of data traffic that they carried began to increase, it became clear that there was a strong justification for introducing data network facilities.

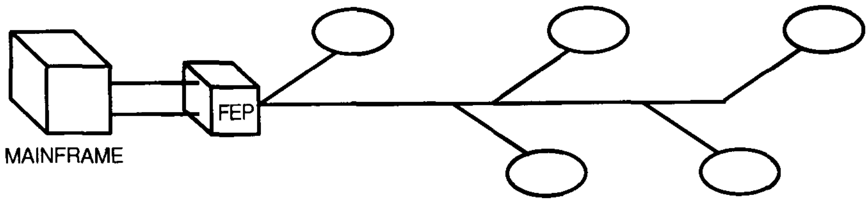


Fig. 4 Main frame linked to terminals by a "front-end processor" (FEP)

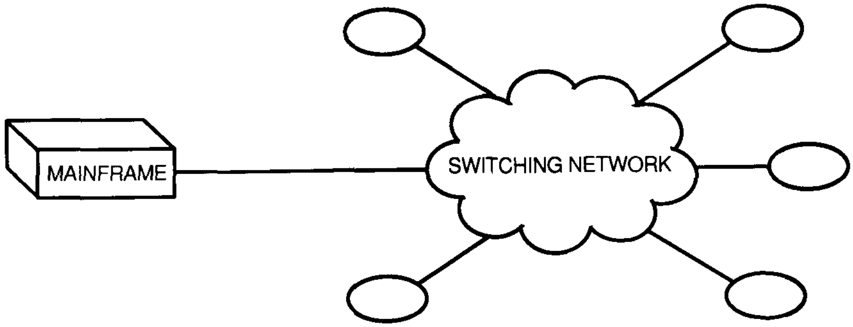


Fig. 5 Main frame linked to terminals via a switching network

3 Wide Area Data Networks

3.1 The Driving Force Behind Wide Area Data Networks

The driving force behind wide area data network development came from the Post Telephone and Telegraph authorities (PTTs) and the private network operating agencies. At that time, most PTTs had a national monopoly to provide common carrier services and they aimed to extend their service to include data networks. Their monopoly did not extend into buildings and so they had little interest in LANs; the LAN story will be unfolded later. The PTTs worked through the CCITT as their international standardisation body.

3.2 The Debate About Circuit Switching and Packet Switching

In the late sixties, the favoured direction was towards the installation of circuit switched data networks, operating like the telephone network but with much lower call establishment times and faster transmission rates.

There were some alternative proposals for packet switching systems, notably from the UK National Physical Laboratory (Donald Davies). These systems require messages to be broken down into maximum length packets before submission to the network. The network routes the packets to the addressed recipient, where the original messages are reassembled. Packets from several users are interlaced within the trunk connections to share channel capacity.

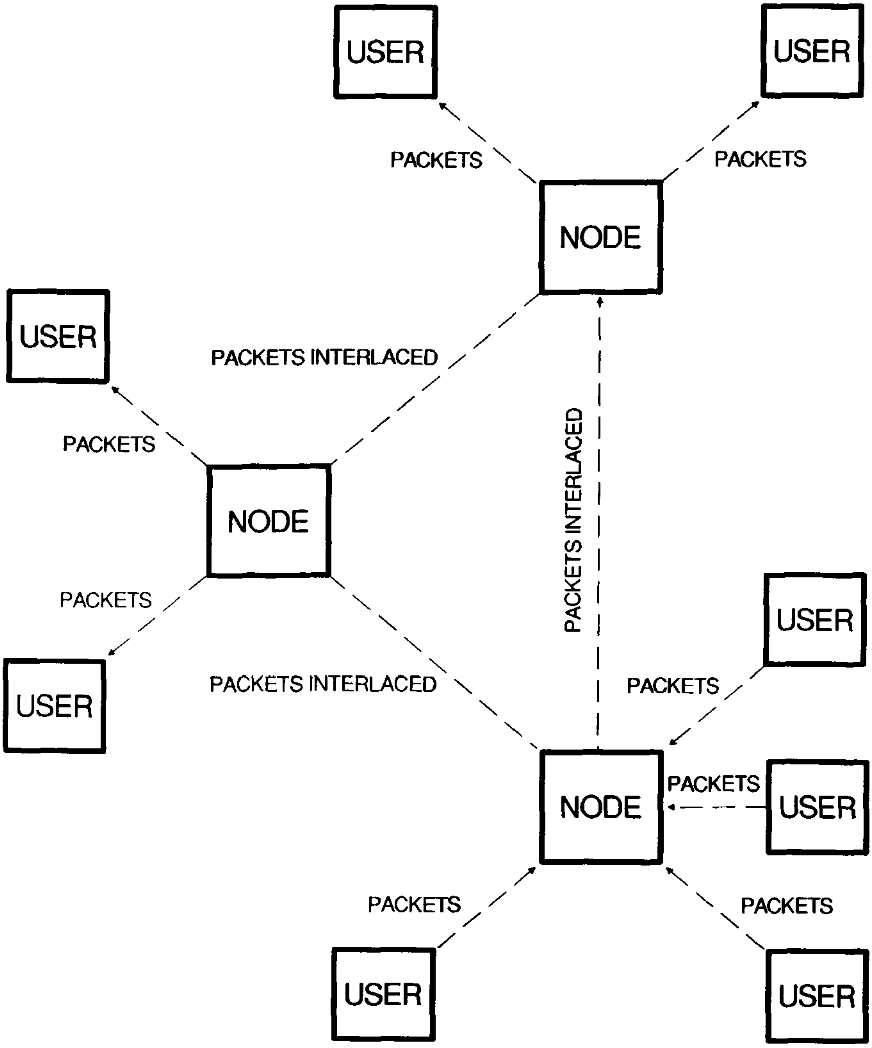


Fig. 6 Groups of users intercommunicating via nodes between which packets are interleaved for transmission

Studies were undertaken to assess the relative merits of the two systems. Large scale integrated circuits were not available at the time and it was concluded that packet switching systems would not be commercially viable.

Circuit switched data networks seemed to be the only solution.

3.3 *Packet Networking Fights Back*

Events moved rapidly. Large Scale Integration (LSI) began to make packet switching networks look more attractive. In 1972 the UK Post Office (now BT) decided to introduce an Experimental Packet Switching System (EPSS) in collaboration with several UK IT suppliers. The transmission protocols were nothing like those which are used today, but the system was made to work successfully and it demonstrated the viability of packet switching.

The plans for a circuit switched data network in the UK were shelved.

3.4 *The Development of Packet Network Standards*

The success of EPSS stimulated the UK into taking an active role in the development of standards for packet switching networks. The ISO had already produced a new data link control standard, known as HDLC, which was ideal as a base for the packet network access protocol set. The CCITT added network control and routing functions to create a packet network access standard (Recommendation X 25). The first version was published in 1978 and confirmed at the next CCITT Plenary in 1980. It has been enhanced over the years but remains the dominant packet network access standard.

X 25 has one subtle difference from the early NPL proposals. They were based on a 'datagram' service which allows packets to be transmitted through the network without pre-establishing that the recipient is able to accept them – now described as *connectionless-mode*. The PTTs decided to offer a service which pre-establishes that the recipient is available before accepting packets for transmission – now referred to as *connection-mode*.

Note that LANs, which were not influenced by the PTTs, normally operate in connectionless-mode.

Most PTTs have now introduced a public X 25 packet network service. Some private carriers have installed equivalent services wherever national regulations permit them to do so. Many private organisations have installed their own corporate service and some share their network with other users.

4 ARPANET and TCP/IP

There was another wide area network influence that must be mentioned. As early as the mid-sixties, USA universities had started to enter into time-sharing arrangements with their colleagues in other universities. At first this was on an ad-hoc basis, but the practice started to grow. Some universities, which were conveniently situated geographically, began to route messages on to other destinations. This grew until the situation became unmanageable and order was restored by introducing the ARPANET research network to route messages within the community. This was done by adding front-end switching nodes, as mentioned earlier, and designing common protocols.

ARPANET has its own set of transmission and routeing protocols, known as TCP/IP, and these have been refined over the years. IP – the Internet Protocol – is a datagram type service. Applications were also evolved to run over TCP/IP, known collectively as the Internet Protocol Suite (IPS). These protocols have also found their way into the private user sector and they are widely used today.

Many users have adapted their TCP/IP systems to operate over X 25 and this provides a key to the coexistence of TCP/IP and OSI protocols in the global environment. The fact that TCP/IP is a datagram service and X 25 is a connection-mode service is no bar to running TCP/IP over X 25.

4.2 What Happened to Circuit Switching?

Scandinavia and Germany went ahead and installed circuit switching services. They were conditioned by the cultural development background in those countries, which leaned heavily towards the evolution of existing telegraph principles. Germany, in particular, were keen to introduce a fast telegraph service, known as *Teletex*, which has not really blossomed (do not confuse Teletex with the TV Teletext services). All the countries which installed a circuit switching service have now installed a parallel X 25 service but circuit switching services still exist and they have many users. However, many circuit switched data network users run X 25 as an end-to-end protocol after the call has been established.

The new Integrated Services Digital Networks operate a form of circuit switching but they also have provision for carrying X 25 packets.

Hence, X 25 will endure as a key protocol for interconnection for many years to come.

5 Local Area Networks

5.1 The Driving Force Behind Local Area Networks

Local Area Networks had quite separate roots. The driving force came from the IT suppliers and was aimed at users who needed a simple cabling system for linking terminals on a single site or campus to the local time-shared processing resources. This was an area where the PTTs had little or no influence.

Many simple systems were developed in the early seventies, mostly based on traditional star connected networks, with multi-drop connections and adaptations of modem drivers. The main breakthrough came in the mid seventies with the announcement of the Xerox *Ethernet* (TM) system. This set the norm for transmission rates at around 10 Mbits/sec for the next decade.

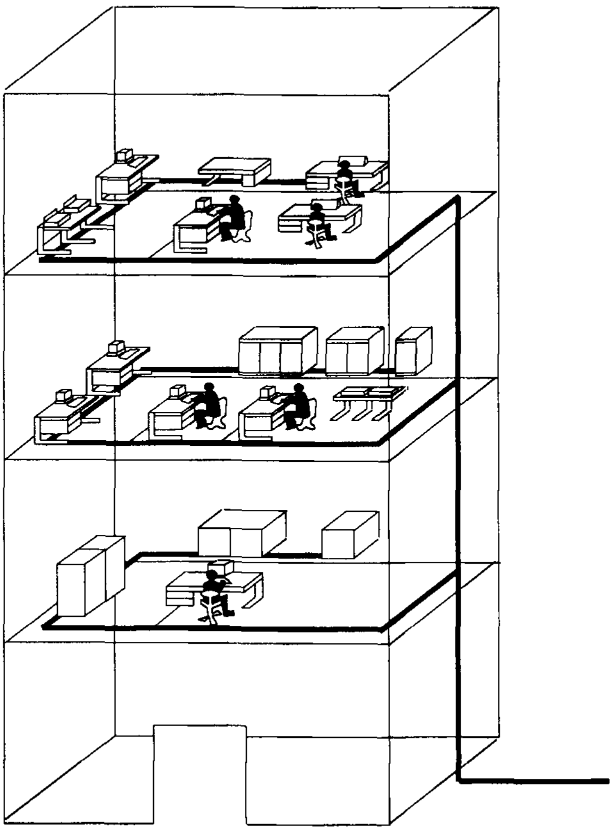


Fig. 7 A local area network in a single building

5.2. The Evolution of Ethernet

Ethernet was an adaptation of a system which was developed at Hawaii University for sharing access to mountain top radio repeaters in the ALOHANET network. The principle is that all stations are free to transmit data whenever the common radio channel is free but they monitor the airwaves to check if any other transmitters start up at the same time. If a 'collision' occurs, all the active stations shut down and try later.

Xerox adapted the ALOHANET principle for use in a multi-access cable network and Ethernet was born. Intel and several other silicon manufacturers invested in integrated circuits and this made the system commercially viable. The technique was documented by the Institution of Electrical and Electronic Engineers committee and the IEEE802.3 Carrier Sense, Multiple Access/Collision Detect (CSMA/CD) standard was produced.

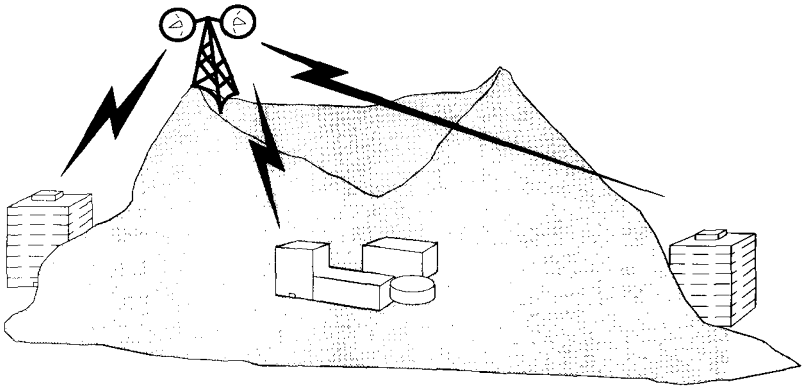


Fig. 8 Illustrating the ALOHANET of the University of Hawaii that used a single radio channel

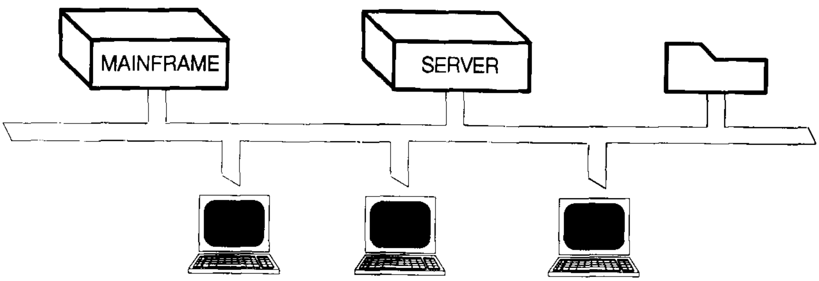


Fig. 9 A local area network (eg an ETHERNET) linking a main frame to one or more servers and a number of terminals

5.3 Other LAN Types

Once the interest of IT suppliers and users was focused on Local Area Networks, other LAN types were standardised.

Token Ring has an IBM pedigree and is aimed at the same applications as CSMA/CD.

Token Bus was developed specifically for factory automation systems and normally works over broadband networks. It was initially sponsored by General Motors.

Slotted Ring was developed from the Cambridge Ring system, which was popular with the UK academic community during the mid seventies. However, its use has declined because it lacked the commercial support enjoyed by Token ring and CSMA/CD.

In 1983 the ISO adopted all the above LAN standards and blended them into the OSI architecture, which means that it doesn't matter to the layers above which LAN type is used. Fitting them into the common LAN/WAN architecture also simplifies the interconnection between LANs and WANs.

5.4 PC LANs

The availability of cheap LAN technology spawned a new industry; the supply of simple resource-sharing PC networks.

Each supplier has developed his own proprietary PC Network Operating System with little attention to open standards. The more reputable PC LAN network suppliers are now beginning to plan support for OSI protocols and to provide migration routes to OSI applications.

Fortunately, fierce competition has ensured that only a few of the proprietary PC networking regimes have a significant installed base and this limits the number for which OSI migration strategies are important.

5.5. Joining LANs and WANs Together

With complex global networks it is a major advantage to design for easy transition between successive LANs and WANs.

LANs and WANs present a similar Data Link Service and this makes it easier to interconnect a LAN system with a WAN system. Gateways which carry out the necessary conversion at the boundary between LANs and WANs have been available for many years but more attention is being given to standardising these functions. The new breed of gateways embrace both the relaying functions, which make the procedure changes, and the routing function, which deals with addressing and route optimisation. They are now referred to under the generic title of *Routers.*

Routers are available which can handle both OSI and TCP/IP protocols and some of the more common proprietary PC network protocols. These are very useful when migrating multi-vendor systems to OSI.

6 Future Evolution

6.1 Integrated Services Digital Networks

Evolution within the wide area has continued towards the introduction of integrated voice and data services. The Integrated Services Digital Network (ISDN) connects transparent high speed channels to subscribers. The wide availability of high speed channels opens up brand new opportunities, such as desk-top conferencing. ISDN also provides a multiplexing path for slow speed control and monitoring devices and this will provide a break out for an entirely new 'Home Electronics' market.

6.2 *Frame Relay and Fast Packet*

The quest for rapid transit through routers has led to new switching techniques which use the simple data link frame addresses for identifying predetermined paths. The address field is the only part of the data link frame which is processed by intermediate routers. All other functions, such as error-checking, are dealt with by the two end systems. This minimises the processing at each router and minimises the switching overhead. Future extensions are planned to cater for more flexible routeing. The technique is referred to as *Frame relay*. It is sometimes referred to as *Fast Packet* but this is really a colloquial term which covers a wide range of techniques for minimising overheads, particularly in relation to transmitting voice and other real time services in packet environments, where it is also referred to as *Asynchronous Transmission Mode* (ATM).

Current Frame Relay systems are extensions of the ISDN data link frame protocol and are designed for use over 64 Kbits/sec and 2.048 Mbits/sec ISDN digital channels which have very low but error rates.

6.3 *Multi-service LANs*

Work in ISO and the IEEE is continuing towards standards for Multi-service LANs which support multi-media services. These LANs run at speeds of at least 100 Mbits/sec over fibre optic links and serve an area of around 40 Km radius. These will carry data, real-time voice, video and other services and they have become known as Metropolitan Area Networks or MANs.

7 Networking Architecture

7.1 *A Structured Model for Networked Systems*

In the early 1970s, the standards organisations began to build a framework for standardisation at all levels, known as the Open Systems Basic Reference Model. This architectural model is now the accepted template for designing integrated business solutions which can operate on a global basis through all kinds of LAN and WAN systems.

The seven layers represent the logical sequence of functions which are carried out when constructing messages for transmission and dismantling them on reception. The electrical conversion of the data is the last function to be performed prior to transmitting the message through the physical network and the conversion back to a digital form is the first function to be performed at the receiver.

The model caters for a wide variety of network environments. How each layer carries out its role in the overall process depends on the demands of the specific environment which can be anything from a cabled local area network to a satellite-connected wide area network.

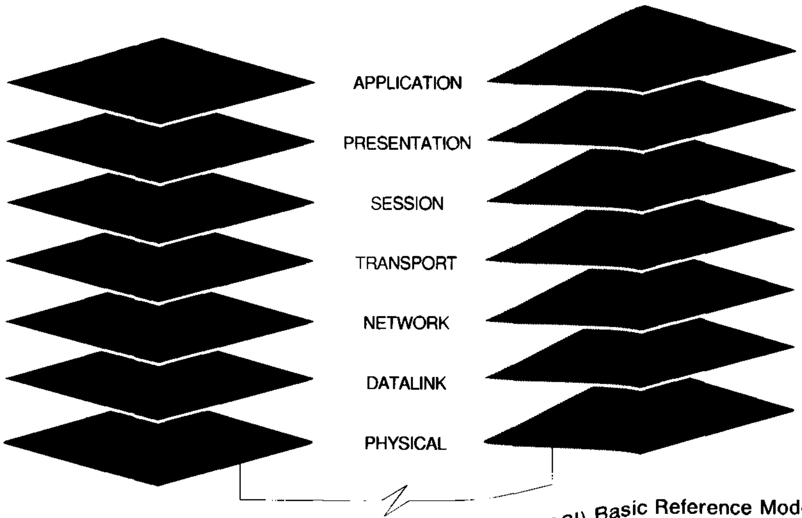


Fig. 10 The 7 layers of the Open Systems Interconnection (OSI) Basic Reference Model

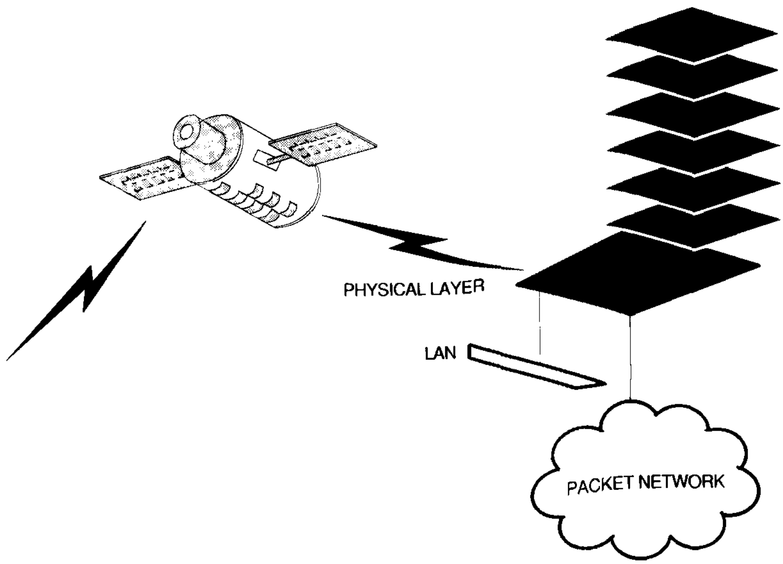


Fig. 11 Illustrating alternative forms of connection at the physical level of the OSI model

7.2 Simplifying the Model

The seven layer model can be simplified by splitting it into two parts with the Transport Layer as the dividing line between two basic sets of functions: Interconnection (the networking functions) and Interworking (the user services which operate over the network).

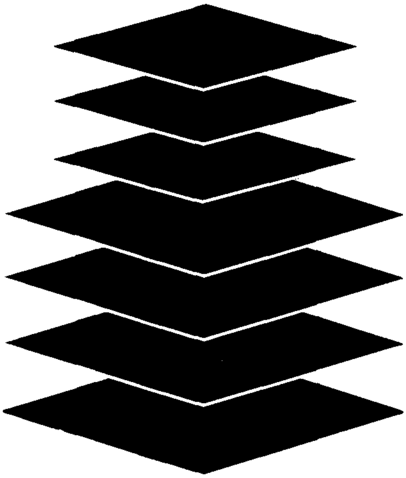


Fig. 12 Illustrating the division in the OSI model between the 4 lower or INTERCONNECTION levels and the 3 upper or INTERWORKING levels

The Interconnection function allows data to move transparently regardless of the physical medium and the route through the network. In other words, any OSI Interworking service will run over any type of LAN or WAN or any mixture of concatenated LANs and WANs.

Management, Directory and Security services were overlaid on the basic two-function model to complete the overall picture. Open networking simply could not be achieved without these common services.

7.3 *The Interworking Services*

Several standards were created for interworking services but the key ones are *Message Handling Services* (MHS), *File Transfer, Access and Management* (FTAM), *Transaction Processing* (OSI-TP) and *Terminal Access Services* (Virtual Terminal). There are more but these form a good portfolio for most business applications and most suppliers conform to this set.

The MHS standard was originally created for inter-personal message transfer (electronic mail) but it was soon extended to embrace *Electronic Data Interchange* (EDI).

FTAM is widely used to control various file movements between systems but it is also used in EDI applications for transferring information such as price files in retail systems.

VTP was developed to allow terminals to access a wide range of services on a number of different systems.

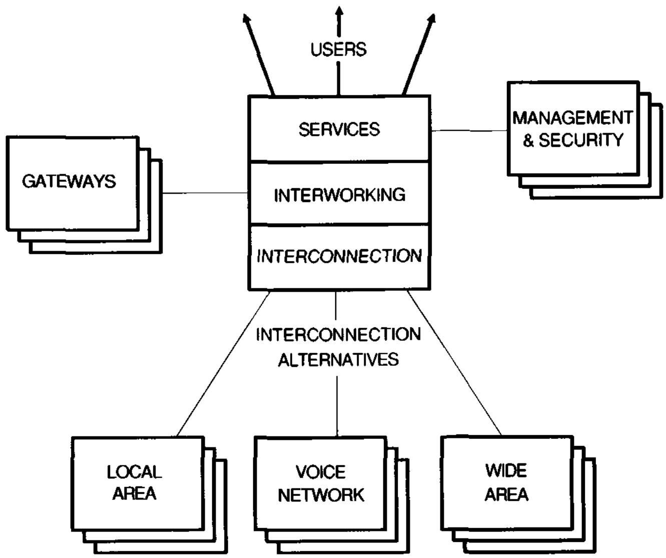


Fig. 13 Illustrating the superimposition of essential management, directory and security services on the OSI model

OSI-TP was developed to allow user applications to share the resources of several systems. It supports information transfer between the various participating components.

7.4 Management, Directory and Security Services

The *Management Service* controls the networks, the systems which connect to them and the software which runs in them. Management services around the network can agree to change the configuration to correct any performance problems.

The *Directory Service* allows users to be accessed through the global network without needing to keep a local record of each user's location address. It works just like a telephone directory.

Security Services control the user's access to the network and the interconnected systems in order to prevent unauthorised access.

7.5 Functional Standards

To minimise the number of permutations of the standards from the seven layers, they are bound into standard combinations within the lower four layers (interconnection) and the upper three layers (interworking) to follow the two function concept which has already been described. The preferred combinations are chosen to meet practical interworking and interconnection requirements and they are known as *Functional Standards* or *Profiles*.

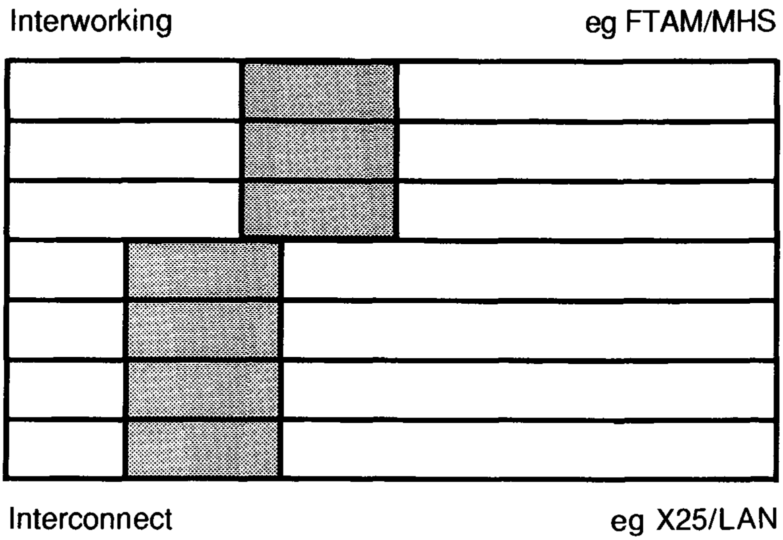


Fig. 14 Illustrating a grouping of "functional" standards in the Interworking and Interconnection levels to minimise the number of practical permutations of standards from the seven layers

Interconnection and Interworking profiles can be fitted together at the Transport Layer to construct any kind of networked solution. The interworking function is unaware of the characteristics of the underlying network.

7.6 Fitting TCP/IP and Other Protocols into the Architecture

De facto systems, like TCP/IP and most of the popular PC network protocols, have a dividing line which is similar to OSI Transport and this helps when blending *de-facto* and proprietary environments with OSI systems. By adopting the OSI transport interface as a common dividing line, users can migrate to an OSI standard network service which supports both OSI and *de facto* application profiles and provides the sound foundation for future global operation.

Similar work within the CCITT is aimed at providing a public integrated services network. This is currently referred to as Broadband ISDN.

7.7 Cordless LANs

Cordless LAN standards are being developed for environments where mobility is a key advantage, such as manufacturing, retail and health. They are sometimes referred to as 'Wireless' LANs. The current focus is on data

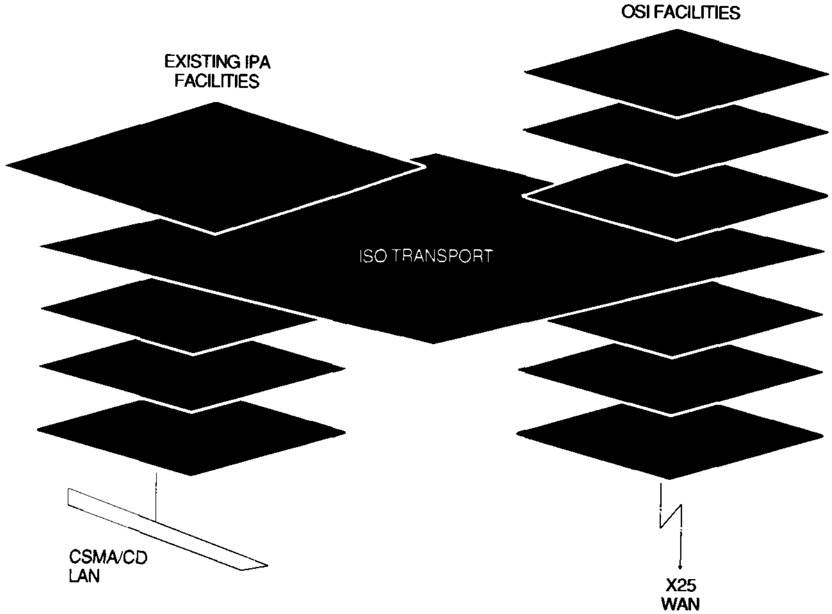


Fig. 15 Illustrating how different Interconnection and Interworking profiles are fitted together at the Transport layer. (IPA = Information Processing Architecture)

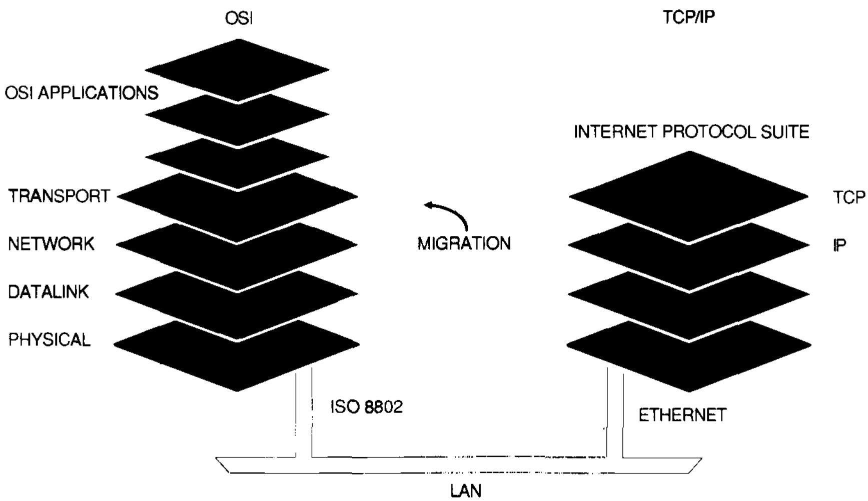


Fig. 16 Showing a migration path from TCP/IP protocol to OSI protocol

applications at modest speeds of around 1–2 Mbits/sec. By the middle of the decade the technology should embrace cellular radio. The transmission rates are expected to increase to 20 Mbits/sec and beyond and they will begin to support multi-media services.

7.8 *Universal Cabling Systems*

The introduction of LANs has focused attention on 'future-proof' cabling systems. There are several LAN cabling systems spanning co-axial cable, unshielded twisted pairs and glass fibre.

These are being rationalised by the creation of an international standard for a universal cabling system, which will be suitable for all the existing LAN systems and the evolution to high speed Multi Service LANs.

8 Adopting an Open Network Strategy

8.1 *Migration Planning*

Many public and private sector enterprises have already adopted an open networking policy. Many others are keen in principle but they need to ensure that their move to open networks is trouble free.

The move presents management with logistical challenges and a high level of risk, unless it is properly managed. The key to successful implementation is a well thought-out migration plan.

8.2 *Practical Considerations*

The principal objective is to make the transition in the most cost-effective manner, and with the minimum disruption to the activities of the organisation concerned.

Start with an examination of the role of the existing distributed information systems which are already in place and consider how they can be brought together into an overall IT commonwealth which includes appropriate business and trading partners.

The plan must ensure that the resulting IT solutions match the future business aspirations of the enterprise and that there is sufficient experience within the organisation to carry it out.

8.3 *Using the Open Network as an Integrating Platform*

An open network provides an integrating platform for sharing information throughout any community of interest but it is only the first step in an evolutionary cycle.

With the integrating platform in place, users can begin to apply their business vision to seeking new ways of exploiting the commonwealth of knowledge which it creates. Really aggressive users will completely rethink their business culture and break out into completely new business dimensions.

New and revolutionary uses of integrated networks provide a real key to global business success.

Further reading

1 *The OSI Handbook*. J. Houldsworth. Published by ICL. Reference G3858, 1991.
2 *The Networking Handbook*. J. Houldsworth. Published by ICL. Reference G3769, 1991.
3 *The Electronic Messaging Handbook*. J. Houldsworth. Published by ICL. Reference G3798, 1991.
4 *Open System LANs and Their Global Interconnection*. J. Houldsworth, A. Flatman, M. Taylor, K. Caves and K. Crook. Butterworth-Heinemann, 450 pp. ISBN 0 7506 1045 X. 1991.
5 *The DISC Guide to Open Systems*. J. Houldsworth. BSI Customer Services – Publications, Milton Keynes, MK14 6LE.

Biography

Jack Houldsworth

Jack Houldsworth has had a distinguished career in information technology design and development. He is currently the ICL Fellow in Open Systems.

He has been involved in standards work since 1966. As chairman of the BSI committee on Data Communications he has led the UK delegation to ISO/IEC JTC1/SC6 meetings since 1973.

His foundation work on the architectural approach to network system design, during the late sixties and early seventies, was adopted by ISO as the basis for the work on Open Systems Interconnection.

Infrastructures of Corporate Networks in the Nineties

Alan Flatman
OPEN *framework* Division, ICL, Kidsgrove, UK
John Lawe
ICL Network Service, Stevenage, UK
Bernard Russell
Northern Telecom Europe, London N11, UK

Abstract

The Nineties will see increasing change in business practices, coupled with the application of distributed IT architectures. The paper examines many of the new and demanding requirements this change will place on the infrastructure of corporate local and wide area networks. Users will need to develop a flexible and open network strategy, selecting the right blend of networking technologies to meet the evolving needs of their organisations.

Typical networks are illustrated and discussed showing how the shape of the corporate network infrastructure is expected to evolve to meet users' future needs.

1 Introduction

Today more than ever before, the IT environment of a business user embraces an assortment of computing and communications systems, often ranging widely in characteristics and based on equipment from different vendors. The associated communications systems have grown to meet the evolving requirements of the user, usually department by department. Local or remote links have been added between these departmental systems until the communications systems spanned first one business location then, as the user recognised the benefits that could result from interconnecting these networks, the communications system has evolved into what is sometimes grandly known as the *corporate network.*

Many organisations now acknowledge that their corporate information network has become a vital resource. The corporate network is now essential to day-to-day operations of the business and affects the way it is conducted.

The corporate information network is now critical to the success of the enterprise, but the IT user is faced with many complex decisions on the best technologies to deploy within it.

As an IT systems vendor, ICL guides customers in the selection of the most appropriate standards and technologies. ICL also seeks to ensure that it has the corresponding products in place to enable customers to implement a strategy which directly matches their business needs.

This does not mean that ICL must support every new technology; it does mean that it needs to support the right set to give customers a competitive advantage. This is achieved through partnerships with the best suppliers of advanced network products.

This paper reviews the IT user and his network, the technology choices facing him and the way corporate networks will evolve through the 1990's.

2 The IT User and the Corporate Network

2.1 *What is a Corporate Network?*

The corporate network, or "enterprise-wide network" as it is often called, can take many shapes and mean different things depending on the nature of the users' business and how it is organised. It can refer to a data network, a voice network or, increasingly, a combination of the two.

To a large multinational corporation the corporate network could be truly global, interconnecting hundreds or even thousands of sites. Such a network would carry voice, data and image and be in use 24 hours a day. As an example ICL's corporate network, ICL-WINS, provides voice, data and video conferencing services linking sites throughout the world. More than 90% of ICL staff can be dialled directly from the desk. The data network offers X.400 mail connectivity for Officepower; more than 85% of ICL staff are connected to the electronic mail network. X.400 gateways are provided to many public services with additional links to other messaging services e.g. fax, telex's Unix mail and Internet. There is a network of LAN to LAN bridges using kilostream links, and a X.25 network which is evolving to replace the bridge network.

Most large businesses operate communication networks at several levels. At *factory or office level* there are likely to be local area networks interconnecting *mainframes, departmental servers, user terminals and PCs.* These business locations will be linked to form a common data network using either point-to-point links or an X.25 network or in some cases both. Again at factory or office level there will be a *PABX and telephone system,* with *telex and facsimile machines,* which will also link business locations. These three functionally separate networks would be carried by private wide area network (WAN) links between business or operational units. These WAN links could thus be seen as a *corporate backbone network* carrying inter-site data and inter-PABX voice traffic over relatively high-capacity privately-leased circuits at 64 Kbits/sec or 2 Mbits/sec.

Inter-company traffic, ordering/invoicing etc. (traditional EDI), E-mail, including in some instances, fully electronic applications for multi-enterprise projects (e.g. in the aerospace industry), will be carried via gateways across the public network or a Value Added Network. In many cases the network grew in a relatively uncontrolled fashion, as new requirements to interconnect sites for either data or voice arose and further links were added. Eventually the user needed a strategy for planning and implementating a true enterprise-wide network to save costs and promote efficiency.

The main options for implementating a corporate network are:

- a *private network* based on dedicated circuits leased from the Telecomms Operators (TOs). This option provides a fixed bandwidth at a predetermined cost.
- a *network using purely public services*, for example X.25 or ISDN. The user is provided with bandwidth *on demand*, which, for occasional use, is the cheapest option. These first two options can be used in combination, the public network being used on demand to supplement the fixed bandwidth available from the private network.
- a *private virtual network* (PVN). A PVN is supplied by a TO and provides a user with managed services rather than just communications paths. Examples of such services are Centrex, X.400 mail and a Frame Relay data service. This last option also provides bandwith on demand [Meers 1992].
- a *Facility Managed network*. In this case a user retains ownership of the networking equipment and contracts with a third party to provide a managed service allowing the user to concentrate on his main business.
- "*Outsourcing*" goes one step further than facility management. In this case a user's networking equipment is sold to a third party reducing the capital invested in the network.

Whatever option is chosen, the costs and risks in implementing an enterprise-wide network are considerable and whole lifecycle costs need to be taken into account.

2.2 Why is a Corporate Network Needed?

A corporate network allows people to communicate more effectively and, for large multinationals, use of Electronic Mail and File Transfer can reduce the effects of time zones on the communications process.

Corporate applications can be supported by an enterprise-wide network, allowing improved supply-chain management, financial reporting and more efficient sharing of central information via central databases.

2.3 *The Role of the Corporate Network*

The role of the Corporate Network is changing. Traditionally the network has supported general administrative functions of the business. Direct revenue earning activity, or industry-specific applications (as distinct from general administration) are forecast to increase both intra- and inter-organisation communications (Inteco Trends, 1990). The operation of the business is changing as many new functions and business processes are now possible by using an integrated, enterprise-wide network. This has been demonstrated in the retail business environment by (Pickworth, 1991) where a network is used to great effect in reducing both costs and time in the merchandising cycle. Enterprise-wide networks also handle information flow to the external business partners, to banks for cash and credit transactions and to suppliers for ordering, accounting etc. As businesses become increasingly global so must IT systems and networks. Turbulence in the business environment will continue, with increased global competition. The number of large global competitors in any business segment will fall with mergers and acquisitions while their size and complexity will increase. (Scott Morton, 1991).

As information is more widely and easily available throughout a business, having the required information in the *right place* at the *right time* is vital to face global competition. Full integration of the IT system improves efficiency of information flow as well as improving efficiency of staff. IT will allow a revolution rather than evolution in business practices.

The change to networked applications means that more information is made available in real time, which can result in an improved image to the customer, more efficient use of valuable resources. It can allow "Just-In-Time" (JIT) ordering systems for better use of working capital (Pickworth, 1991).

Increased emphasis on the ease of use of systems, the Human Computer Interface (HCI) , has allowed access by all sectors of the workforce and led to the "PC on a desk". Adoption of "Multimedia" applications will allow presentation of fuller information through use of sound and moving image.

A most important change in IT architecture has been the move towards Open Systems. Latterly Open Systems Interconnection (OSI) has changed from being a good idea, to seeming very desirable, or even in several market segments to being mandatory, notably in Central Government through GOSIP. Currently a mix of proprietary, de facto and Open standards is in use. Only OSI will allow true multi-vendor interworking, (Houldsworth et al., 1991).

Use of client-server architecture exploits cheap processing power in workstations by placing the application as close to the end-user as possible. The client/server communication traffic generated by these locally-based applications frequently requires interconnection of systems supplied by different vendors, which must therefor conform to common standards. Adoption of client-server architectures is forecast to increase rapidly in the early nineties (Gartner, 1990).

Technology now allows applications to be more widely distributed. In the 1960's and 1970's applications ran on a single mainframe with users gaining access to the service via dumb terminals star-connected to the mainframe (Houldsworth, 1992). In the 1980's with the advent of Local Area Networks (LANs) local distribution of service was possible. For remote users access to the central system was by leased or dial-up communications lines. As more applications are delivered via a LAN, high-performance LANs are demanded—hence the FDDI backbone.

In effect the role of the network is changing and the requirements placed on it become more stringent. To create an enterprise-wide network an integration of WAN and LAN technologies is required, which is a target for the 1990's. The high bandwidth available over a LAN has allowed development of distributed applications and client-server architectures. True integration of LANs and WANs will mean that distributed applications and client-server architectures available locally on an individual LAN, are made available at other locations via the WAN with the same short access and response times. Application platforms can be sited and the applications themselves can be distributed, wherever convenient, not necessarily where the users are. Multiple instances of the same application need no longer be provided for access reasons. The users need not be aware of where the applications are, or of the mechanisms needed to access them.

To sum up, the Corporate Network is one of the key assets of a business. The potential of that network must be exploited to the full to meet immediate challenges while laying a firm foundation for future enhancement. A flexible networking architecture must be developed that is capable of responding to the changes that will occur in the company, its business and thus its future IT requirements.

Combining previously separate services (IT, telecomms and building management) on one common infrastructure offers very real advantages to the network provider. There is only one network to install, maintain and manage. A network with its multiple paths between nodes is inherently robust; in a well-structured network the failure of any individual node should affect only users connected directly to it. If an even more robust system is required each user can be connected to several nodes independently. Flexibility also in meeting future needs can maintained in terms of bandwidth allocation allowing the traffic mix to change with time.

3 Technical Characteristics of Networks

When planning the design and implementation of a network it is obviously vital that appropriate technologies are chosen to meet the business requirements of the user. Key network characteristics to be considered in matching technologies to the requirements include:

Robustness — If the business depends on the network, availability and resilience against component failures are essential.

Open-ness — How to allow for multi-vendor support and future flexibility.

Security — Sensitive business data now being carried across the network.

Multiple Services — Is the same network required to carry voice, data, image, telemetry etc?

Costs — First costs should be balanced against whole life-cycle costs.

Geography — The network needs to be resilient to changes in geographical requirements, ranging from movement of individuals between sites to the integration of multiple ,existing large-scale networks in the case of company mergers.

Performance — Parameters that need to be estimated and specified will include:

Overall throughput, to meet the load requirements.
Instantaneous throughout, to cope with varying traffic patterns.
End-to-end delay, important for interactive use.
Delay variability, important for voice and video traffic.
Residual error rate, important for data.

Manageability — To satisfy all the above criteria the network must be manageable in a cost effective manner.

4 Summary of Technology

Technologies becoming increasingly important to planners of IT networks include the following. These are described in more detail in companion papers.

Frame Relay — A packet-switching technology in high performance data systems. Principally used as a LAN-to-LAN interconnection in conjunction with "router" technology. (Meers, 1992)

Cell Relay — Also known as Asynchronous Transfer Mode (ATM), it is similar to Frame Relay but uses small, fixed length cells to transport information. Applicable to voice, data and video. (Meers, 1992)

FDDI — Fibre Distributed Data Interface is a standard for a 100 Mbits/sec fibre optic network using token-passing ring technology. Used for high bandwidth backbones. FDDI-2 is a hybrid technology designed to carry both voice and data. (Taylor, 1992)

ISDN — An Integrated Services Digital Network conforming to a set of CCITT standards for voice, data and video.

ISPBX	An Integrated Services PBX will allow switching of integrated voice and data traffic to ISDN workstations.
MANs	Metropolitan-Area Networks provide LAN-to-LAN switched connections covering an area up to 50 km in diameter. IEEE standard 802.6 applies.
IVD LAN	Integrated Voice and Data Lans will allow transmission of data, image, real time voice and video via a single LAN. IEEE standard 802.9 applies.
Wireless Networks	Wireless LANs behave as ordinary LANs but require no wired connections. Wireless Public Access Data services can provide National packet switched data facilities using cellular technology. (Flatman, 1992)
Low Bit Rate Voice	A scheme for compressing voice traffic to allow several voice channels to be transmitted over a 64 Kbits/sec circuit.

More detailed information on the above technologies can be found in "Open System LANs and their Global Interconnection" by (Houldsworth et al., 1991).

5 Networks in 1992 and the Future

The figures reproduced below are intended to give typical pictures of corporate IT networks now and in the future. However it mst be stressed that these must not be seen as universal. Each user will have a different mix of requirements, and therefore will seek a design of networks tailored to both his immediate and his long term needs, fitting the design to his business, not the other way round.

5.1 Typical Infrastructure in 1992

The scenario shown in Figure 1 is characterised by the use of private circuit-switched or point-to-point WAN links for both voice and data between the larger corporate sites. Their high performance but low connectivity needs are met by circuits operating at from 64 Kbits/sec up to 2 Mbits/sec with multiple 2 Mbits/sec circuits where required. The PSTN, in conjunction with public and private packet-switched data networks, connects the smaller sites whose medium performance but high connectivity needs are met by circuits operating from 9.6 Kbits/sec up to 64 Kbits/sec.

The private WAN is implemented as a simple mesh of point-to-point leased digital circuits or as a multiplexer network. Line rental is a major element of network expenditure. Multiplexers allow voice, data and video traffic to share the expensive point-to-point bandwidth. Proprietary techniques based on Cell Relay or Frame Relay are emerging, which will allow optimum use of the entire bandwidth, instead of a fixed partitioning of the bandwidth

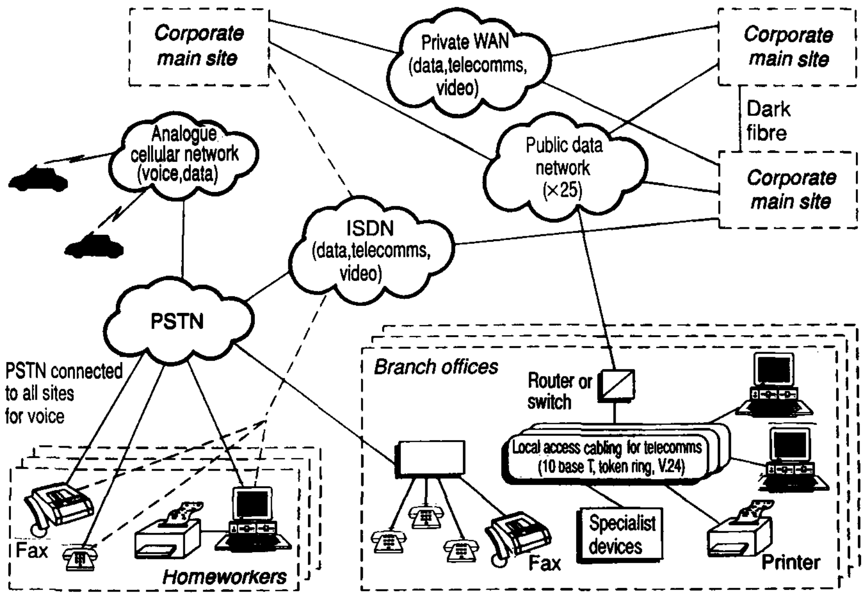


Fig. 1 Overview Corporate Network Infrastructure for 1992

between voice and data. Use of low bit-rate voice can make a 64 Kbits/sec circuit economical by allowing simultaneous voice and data channels.

Higher connectivity configurations, including Value Added Networks (VANs) provided by third parties, employ a combination of X.25 packet-switching and Frame Relay. X.25 is used to meet medium performance requirements with access rates typically up to 256 Kbits/sec. These public X.25 services facilitate inter-company electronic communications such as E-mail and EDI, although most inter-company communication is currently by voice. Where higher performance is required Frame Relay in conjunction with routers may be used for LAN-to-LAN interconnection with typically up to 2 Mbits/sec access.

The split between public and private WAN varies from country to country, dependent on differing tariffing policies. Private WAN networks are dominant in the UK, often extending to private links to Branch Offices using multiplexers and/or X.25. In countries such as France and Germany the public packet-switched or circuit-switched networks are more widely used.

In the UK ISDN is used currently only in niche areas and where it can support existing technologies, such as standby facilities for leased 64 Kbits/sec circuits or provide 64 Kbits/sec bandwidth to homeworkers.

There is no significant *integration* of transmission of voice and data other than co-existence within a private WAN multiplexer network.

Local area networks are increasingly being based on structured cabling, with multimode optical fibre often providing inter- and intra-building backbones. Twisted copper pairs, primarily unscreened, provides a popular mode of connection to the desktop.

FDDI is replacing CSMA/CD and Token Ring for the backbone network higher bandwidths to be carried between buildings or across a campus. 10BaseT and Token Ring are more commonly used for desktop attachment, with 10BaseT dominating by virtue of its maturity and low cost. V.24 connection to the desktop is being replaced by 10BaseT. Voice and data still remain separate.

A router (or internetwork gateway) is rapidly becoming the standard method of interconnection between LAN and WAN domains. Routers support Open standards on the LAN and other support both for differing LAN media and a variety of internetworking protocols. On the WAN, access to X.25, point-to-point and Frame Relay networks are all commonly supported. The router is used as major point of separation between the two domains, local and wide area, allowing the technologies to develop independently.

Bridges and routers are increasingly being incorporated into an *intelligent hub*, allowing the local area topology to be implemented using hubs with integrated support for the WAN.

Leading edge technologies available but not yet in common use include:

- ISDN
- Low bit rate voice in the private WAN
- Frame Relay in the private WAN
- FDDI backbone in the local area network
- Integrated Services PBX
- Video-Conferencing
- Use of dedicated fibre links to interconnect nearby main sites.

Mature technologies, at or near the end of their development, include:

- X.25 packet switched networks. The technology is now reaching its performance limits, although it is expected to have a significant role for at least a further 5 years.
- V.24 local connections. These are being replaced by LANs although V.24 will persist for low function terminal connections via structured cabling.
- TDM multiplexers. These are being superseded by proprietary packet-switching mechanisms based on Cell and Frame Relay techniques to multiplex both voice and data traffic.

- PSTN will remain in use for data, where it is cost effective, but ISDN and digital point-to-point working will supersede it where higher performance is needed.

5.2 Typical Corporate Network in 1995

Figure 2 illustrates continuing use of private WANs, complemented by a new generation of public WAN services which, between them, offer a wide range of performance and connectivity options.

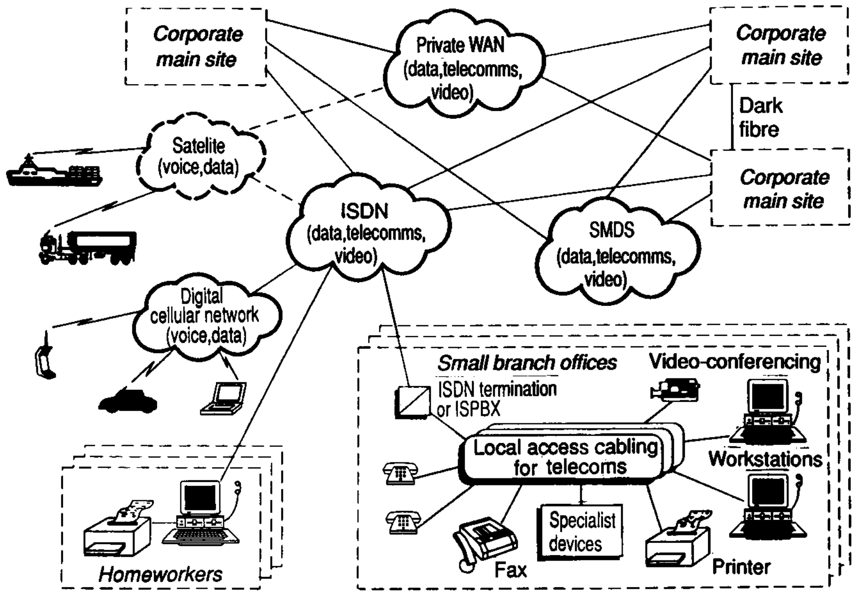


Fig. 2 Overview Corporate Network Infrastructure for 1995

The private WAN represents an evolution from the current strategy consolidating and converging the various elements. To keep network running costs down efficient use of bandwidth will be a major requirement.

The underlying point-to-point links are expected to be in the 2 Mbits/sec to 34 Mbits/sec range, using SDH (Synchronous Digital Hierarchy) which allows Telecom Operators to provide services on a flexible and efficient basis [Meers 1992]. The private WAN switching represents a convergence of the traditionally separate multiplexer and X.25 switching technologies, and handles voice, data and video equally efficiently. It is anticipated that Frame Relay will replace X.25 as the principal technology for switching data between main sites (LAN-to-LAN), with growing support for data rates up to 34 Mbits/sec. Cell relay (based on ATM Asynchronous Transfer Mode principles) is expected to provide the backbone multiplexing and switching over the available point-to-point bandwidth.

Telecomm Operators are expected to provide Private Virtual Networks for voice, data and video over SDH.

Structured cabling is forecast to be commonplace, with optical fibre continuing to form the backbone, with twisted copper pair to the desktop.

FDDI is expected to be well established as a building and campus backbone, with multiple FDDI rings providing the very high aggregate bandwidths required by some organisations for high definition image and multimedia, etc. There is expected to be an increasing trend for Corporate and Departmental Servers to be attached directly to the FDDI backbone.

10BaseT and Token Ring are expected to become commonplace methods for feeding the desktop, having displaced a large proportion of the V.24 connections. *Dedicated* CSMA/CD or Token Ring, (allowing one station full access to the available LAN bandwidth), via intelligent hubs offers a cost-effective alternative for providing high bandwidth to top-end graphics or high definition image workstation. This technique may also be used to provide secure connections to workstations where appropriate.

The ISPBX is anticipated to be increasingly prominent, providing intelligent switching for local telecomms services. ISBN BRI and $n \times 64$ Kbits/sec services will be provided to the desktop, particularly for the small office.

The debate over public versus private provision of the network is expected to continue, the choice being between the Private Network, Public Network, Private Virtual Network, Facility Managed network or Outsourced.

The corporate network will be flexible and operate worldwide, it will support business trends by freeing workers to work from anywhere, while keeping in contact with corporate and departmental information.

6 Conclusions

6.1 The optimum strategy for building a corporate IT network will be unique to each business. The choices made will be determined not solely by the available technology nor by its cost, but primarily by the special requirements of that business and of course its past history in networking.

6.2 The market is already highly complex with stiff competition to sell or rent products and services to user businesses from equipment and software suppliers, telecommunication operators and a multitude of providers of specialist value-added networks and services. The key underlying conflict is between a DIY and a "package" approach. TOs and speciality houses are providing an increasing range of network services likely to appeal to medium and smaller businesses who may be unable to face devloping and planning corporate networks for themselves.

6.3 The largest businesses, seeing their competitiveness and even their very existence to be increasingly dependent on their corporate IT networks, will continue to retain control by evolving and operating unique networks of their own.

References

FLATMAN, A., WEISER, P. Wireless Networks, *ICL Tech J.* Vol. 8, no. 2, pp. 242–271, 1992.

GARTNER, Enterprise Network Strategies, Sept 1990, (*ENS*: R-304-105).

HOULDSWORTH, J., FLATMAN, A., TAYLOR, M., CAVES, K., CROOK, K. Open System LANs and their Global Interconnection, *Butterworth Heinemann*, ISBN 0 7506 1045 X.

HOULDSWORTH, J. Open Networks – The Key to global success, *ICL Tech. J.* Vol. 8, no. 2, pp. 179–197, 1992.

Inteco Trends 1990.

MEERS, J. Broadband Networking, *ICL Tech. J.* Vol. 8, no. 2, pp. 210–224, 1992.

ORANGE et al. ISDN Feature, ICL Tech. J Vol. 7 no. 3 pp. 451–511.

PICKWORTH, I. Experience of Managing Data Flows in Distributed Computing in Retail Businesses, *ICL Tech. J.* Vol. 7 no. 4 pp, 718–731.

SCOTT MORTON, M.S. (Ed) *The Corporation of the 1900', – Information Technology and Organisational Transformation*, OUP (NY) ISBN 0-19-506 358-9 1991.

TAYLOR, M. FDDI – The High Speed Network for the Nineties, *ICL Tech. J.* Vol. 8, no. 2, pp. 225–241, 1992.

Biographies

Alan Flatman

Alan Flatman is LAN Architecture & Conformance Manager within ICL's OPEN*framework* Division. He is a Fellow of the Institution of Electrical Engineers, a Chartered Engineer and PhD in Electronics & Computing, and has been deeply involved in the development and standardisation of LANs since 1980. He is responsible for ICL's LAN Technology and Strategy.

John Lawe

John Lawe is Technology Consultant for ICL Network Services. A BSc graduate in Electronics, he has been involved in the development of networking within ICL for the last 23 years.

Bernard Russell

Bernard Russell was Technical Strategy Manager for WANs for ICL Network Systems. A BSc graduate in Physics, he has been closely involved in marketing and technical strategies for ICL. He has recently joined Northern Telecom Europe as manager of SDH Management Systems.

Broadband Networking

John Meers
Associated Services Division, ICL, Stevenage, UK.

Abstract

Data networking now plays a key part in the conduct of many business operations. As business needs change, so the demands on the network change. Current business requirements are met by packet networks and local area networks, but new applications using graphics and other techniques require greater bandwidth than these can offer.

Broadband techniques offering throughput of 100 Mbps and higher are becoming available to satisfy these new applications. At present there are a range of competing products and standards which have evolved from existing network techniques, but convergence towards a new type of service based on flexible allocation of capacity is beginning to emerge.

1 Introduction

Data networking is now almost as much a part of todays business scene as the telephone network. Many businesses are as dependant on their data network for the day to day operation of the company as they are on their telephones. However, the telephone network uses well established and stable technology, whilst the technology used in data networking is changing very rapidly.

Todays data networks are based on packet switching and local area networking technology. This paper examines how the technology is moving towards a new range of services offering much greater bandwidth and flexibility to users both of private and public networks.

The paper describes in outline the emerging broadband technologies ATM, Broadband ISDN, Frame Relay, FDDI and others. It reviews development in the standards for these technologies and illustrates how business demands and broadband networking technology are proceeding together to generate new opportunities for the use of data networks.

2 What is Broadband

Broadband and Baseband are terms which are used in Local Area Networking (LAN). Baseband describes the technique of using a single portion of the spectrum from DC typically up to 50 MHz. It is the technique used for most LAN products today. Broadband is a technique of splitting up the frequency, which is sent down a coaxial cable, to support several separate information channels, typically for a baseband LAN plus other services such as video.

Although the term 'broadband' has this specialised LAN use, it is also used more generally for high speed, multi-channel transmission. With the advent of new high speed transmission and networking techniques, this wider use of the term is becoming more common. In this paper the wider definition of broadband is used.

Broadband, used to define high speed transmission and networking, is poorly defined, in fact the definition can be considered to change with time. The term is most often used to refer to networking resources and lines working at higher speeds than those normally available. In the days when most lines operated at 300 bps or 1200 bps then any use of multiplexed 64 Kbps lines (or higher speeds) would be considered as broadband networking. Currently leased line speeds of 64 Kbps and 2 Mbps are common, and for the purposes of this article broadband networking is defined as a service which will provide switched connections at 2 Mbps and above.

In the future these speeds may be common, and not considered as 'broadband' at all. It is interesting to look at the growth of line speeds over the last 50 years or so. In the early years speeds of 50 baud for telex were the limit of technology; even in the sixties 75 baud telex machines were not widespread and 300 bps was considered a high speed link. The use of modems over the telephone network opened the way for higher speeds, and data rates slowly grew from 300 bps through 1200 bps and 2400 bps to 9600 bps. Today speeds of 64 Kbps and 2 Mbps are widely used, and ISDN is providing a switched 64 Kbps service.

Services being developed will operate at higher speeds than this. FDDI operates at 100 Mbps and services using 600 Mbps are being developed. Figure 1 shows the growth of line speeds over this period.

3 Today's Networks

3.1 Private Networks

The technologies used in todays networks are Local Area Networks (LAN) and Wide Area Networks (WAN) or even a simple fixed network of leased lines or multiplexers. A LAN may use transmission speeds of 10 Mbps within a building, but typically only 64 Kbps between buildings. Wide Area Net-

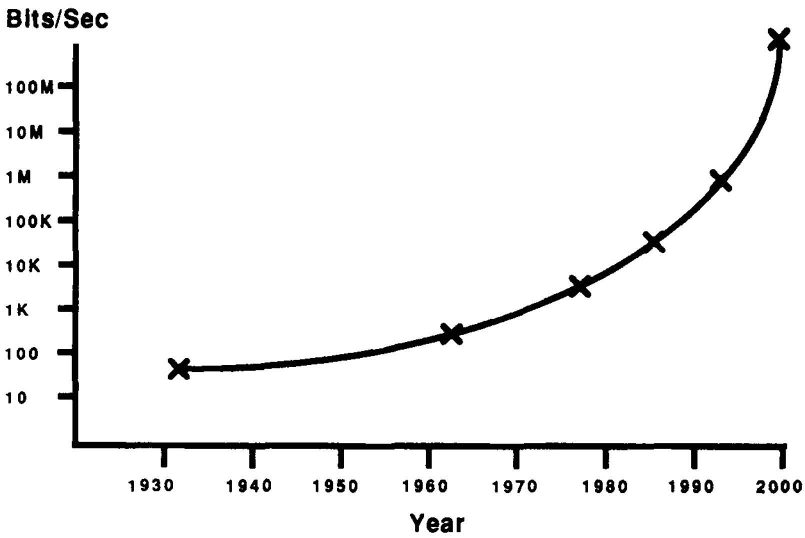


Fig. 1 Growth of Data Speeds

works can be based on modem links over the phone network, or on packet switching based on the CCITT X 25 standards. The line speeds used in such a WAN are usually 9600 bps or even lower, although 64 Kbps is common for trunk links in an X 25 network.

LAN and WAN networks provide switched connections to their users as is shown in Figure 2. In addition an underlying transmission network is needed to provide data links between nodes within the networks. This transmission network provides fixed links between the switching nodes of private data networks and is also often used for digital speech circuits in a private voice network.

A typical transmission network would be based on multiplexers connected by leased lines operating at 2 Mbps and providing 64 Kbps links.

This situation is beginning to change as users demand more complex and faster services. New standards and products are appearing which provide considerably greater bandwidth, and the distinction between a fixed transmission network and a switched data network is becoming blurred.

3.2 Public Networks

The situation with todays public networks is very similar to that for private networks. Private circuits operating at speeds of 64 Kkbps or 2 Mbps are obtainable. In addition public packet switched networks based on CCITT X 25 are available, and ISDN networks supporting 64 Kbps switched links for voice and data are becoming available.

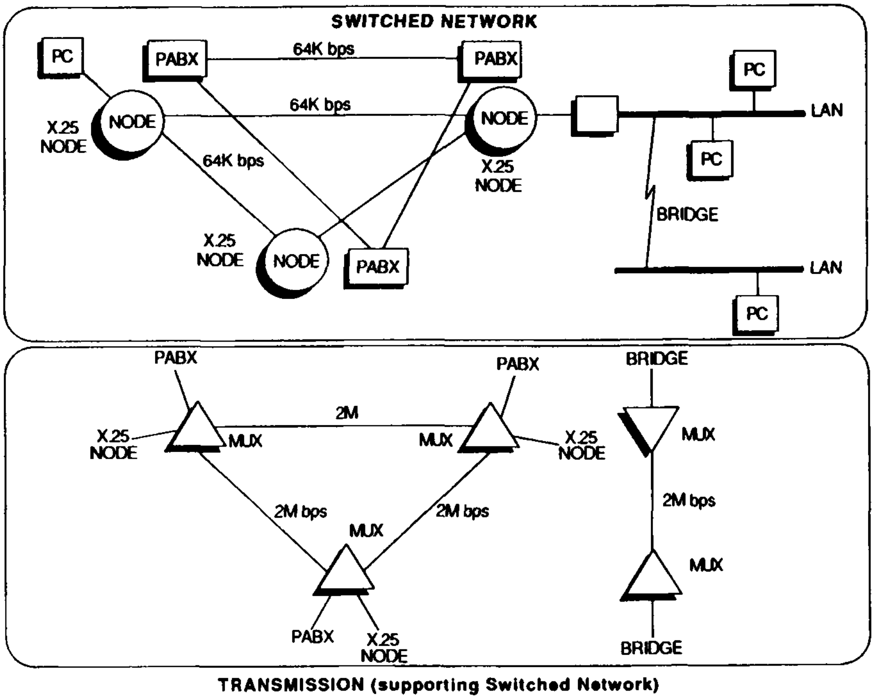


Fig. 2 Today's Networks

4 Business Pressures

4.1 General

Today, networking is used within a company primarily to provide access from a single terminal on a user's desk to a range of applications. This simple approach is changing as users realise that networking can be used as a means of improving their business performance. Two key areas are:

- *Communication with Other Companies* – People become familiar with the use of networks to access or exchange data within their own company. It is a natural step to want to exchange information with business partners as well as within the company, after all the telephone network allows worldwide access for voice communications.
- *Access to More Applications* – Simple networking lets users connect their terminals to remote applications. This is a big step forward, but again it is natural for users to want to extend the range of applications to which they can connect. Increasingly these applications will need more than a simple text input. For example computer aided design (CAD) applications will need very high data rates to operate at acceptable speeds over a data network.

The combination of these business pressures is leading to a new generation of broadband networks for communication within a company, and for inter company communication.

A few more specific examples of how broadband networking may be used are illustrated below.

4.2 Computer Aided Design

Most CAD systems are based on a high speed LAN with workstations accessing a shared server. With networking why do the workstations need to be on the same LAN? The answer is that the high resolution CAD screens need a lot of data from the server. The LAN speed of 10 Mbps is fast enough to give response times acceptable to the user, but network links from the LAN are likely to be only 64 Kbps, or even 9600 bps. In that case the system would be so slow as to be unusable.

If broadband networking is used, with speeds of 100 Mbps or higher, then the CAD workstation can be on another LAN in the building, or even on a LAN in a different office.

4.3 Video Conference

The new ISDN services, providing a 64 Kbps link, allow users to exchange data during a phone call. This procedure can be used for slow scan TV to provide a 'picture phone' showing the person at the far end. Alternatively it can be used to display pictures or documents at both ends of the connection to allow interactive changes during a call with both users seeing any changes (Fuller, 1991).

Current picture phones using 64 Kbps, provide a rather small, low resolution picture. Broadband networking would allow more bandwidth and better picture quality, which could make this type of service more attractive.

The picture quality of personal computer displays is increasing with each new range of computers. If people are looking at pictures at each end of a network link, then they will expect the same high resolution. They will also expect any new pictures to be displayed at both ends with minimal delay. To meet these requirements the higher data rates possible via broadband will be of considerable benefit.

4.4 Database Access

A common use of networking is to access a remote database and extract information. Today most people use terminal emulation and retrieve text information from the database.

Two features are evolving which improve the usability of such access to databases, but which will need higher bandwidths.

- Display of information in the form of pictures of graphs can provide the information in a more friendly form than just text, but requires much more data to be transferred.
- Rather than using terminal emulation, a database product in the users own personal computer may be used. This can then extract the information from the distant database and display it for the user. To do this, the two databases need to 'talk' to each other using protocols such as Structured Query Language (SQL). This uses lots of messages between the two databases to get the information; to give an acceptably fast answer, a high data rate is needed.

4.5 Distributed Computing

As personal and departmental computers become more powerful, people use them directly for many computing tasks, so that the need to connect a terminal to a remote computer system via a network diminishes. Applications are being developed which provide a user interface, and perform part of the processing, on a local machine, but use the network to extract needed information from distant machines. With todays network speeds it is easy to tell if the computing task is local or involves 'networking', since any significant use of networking will cause unacceptably long response times and irritation to the user.

The higher speed available via broadband can hide the use of the network from the user.

4.6 Cable TV

Another area where broadband is being used is in the transmission of television programmes. Modern Cable TV is moving from transmission of analogue signals towards digital. This provides a better quality service and the potential for transmission of a wider range of programmes. Provision of digital TV to the home will need the support of a high speed digital network. This use of broadband is still some time in the future.

4.7 New Uses

As well as the business applications considered above, new uses will emerge. One possibility is to use broadband networking to 'buy' TV programmes. A user could use the network to call a central point, then use the high speed of the network to download a programme to a video recorder for later viewing.

5 Broadband – Tomorrows Networking

5.1 *Evolution and Convergence*

Broadband technologies and products are providing new services to users, but they are in many ways, an evolution from the current network technologies.

5.1.1 Better X 25 One route to broadband is via packet switching. Fast packet switches, can transfer fixed size packets at rates of the order of 1,000,000 packets per second. Suppliers of packet networks often offer this type of product at the top end of their range, either to be used as a fast central core in a large packet network, as an alternative to "routers" for interconnection of LANs, or for other high speed services.

5.1.2 Better LANs Broadband is also evolving from local area network (LAN) technologies. New high speed LAN standards such as FDDI, working at 100 Mbps, are emerging. Most LAN suppliers will now offer FDDI as part of their product range. It is often used as a backbone connecting existing LANs together within a building, or group of buildings (a 'campus').

Routers, bridges and other devices are used today to link LAN's together into larger networks. In general, todays links operate only at speeds of 64 Kbps or 2 Mbps. This is adequate for todays needs, but is not fast enough for many of the emerging business needs.

5.1.3 Fast Configuration Multiplexer A third evolutionary path for broadband starts from multiplexing. Current multiplexers operate at higher speeds than networks, but provide fixed connections. Configuration of the multiplexers is done by a network administrator from a management station. Changes are relatively slow and infrequent, and done only as the network configuration changes.

Several suppliers are offering products based on multiplexers, but with improved configuration software so that users can set up connections themselves. These are often referred to as 'fast configuration multiplexers'. They provide a switched, high speed connection and make the boundary between networks and multiplexers somewhat blurred.

Existing multiplexers available to users, operate at 64 Kbps or 2 Mbps, but higher order multiplexers are used by PTTs and standards for these are also leading towards broadband working.

5.1.4 Broadband ISDN ISDN is a connection-based switched network operating at 64 Kbps. Work on standards for a higher speed network, Broadband ISDN, is also in progress.

5.1.5 Cable TV The above routes towards broadband have been from data transmission. In addition to this Cable TV companies and standard bodies are involved because of the use of digital transmission for TV distribution.

5.1.6 Convergence As well as these evolutionary trends from existing products, convergence towards a common broadband service is evident. A fast packet switch offers a service which is very similar to that offered by a configurable multiplexer. To interconnect LANs a user can choose in many cases between routers, an FDDI backbone or a fast packet switch. A number of general trends for broadband are becoming evident.

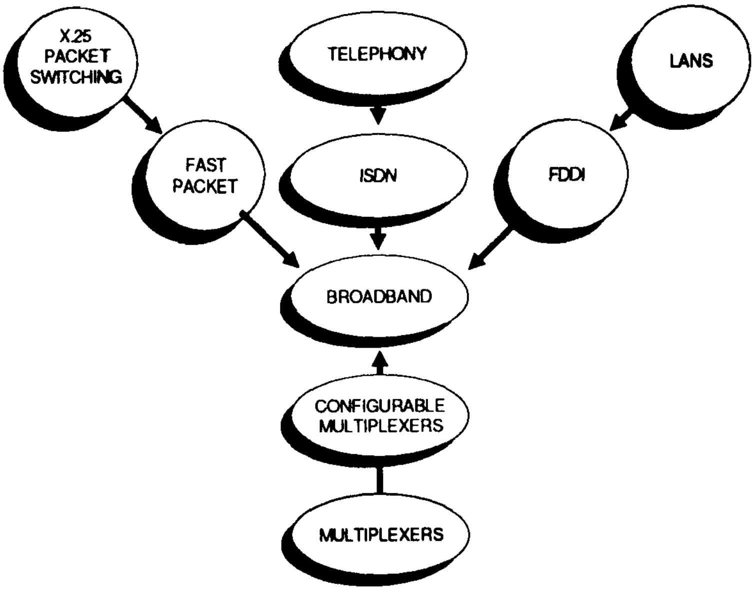


Fig. 3 Evolution and Convergence

5.2 General Trends

5.2.1 Demand Bandwidth One trend which is becoming apparent is that a flexible allocation of bandwidth. Todays networks provide a user with access to the network at a fixed speed. Circuit-switched networks provide the user with capacity to support this speed for the duration of the call even if the user is not sending anything.

Packet-switched networks go some way towards using capacity more efficiently in that the network will only transmit packets when there is data to be sent, although the users link to the nearest packet node is not shared.

Local area networks go even further since the connected PC or other device, only uses the LAN when it has data to send.

The trend in broadband working is towards flexibility in the allocation of bandwidth. In the shorter term this will allow networks to control the capacity of their trunks. Rather than a fixed 64 Kbps or 2 Mbps link, they can request the capacity they need, perhaps 10 Mbps if several users are simultaneously working on CAD over the link, perhaps dropping the link completely overnight if there is no traffic. In the longer term this demand bandwidth will extend to users with applications requesting bursts of capacity for short periods, perhaps to transfer a picture from one PC to another.

5.2.2 Baseband Speeds Another general trend is a demand for higher speeds by services using networks. A human can type text only rather slowly, but can read text somewhat faster. If this was all that was needed then current speeds of 9600 bps would be adequate, and 64 Kbps more than enough. However if pictorial information is sent, then a person can absorb information at a much higher rate. Consider a possible application where someone is using a network to select wallpaper (or any other similar graphic information). They get a picture on their screen, rapidly decide they don't like it and scan to the next.

A further example would be for database enquiries. A user might be using his PC to access information. Much of the needed information may be held on distant computer systems. To satisfy a particular request, the application programme in the PC may need to use a network to get information from several sources and select and collate it to produce the desired answer. To do this without causing a long delay will need very fast call setup and transmission of data over the network.

5.2.3 Packetised Voice So far the discussion of broadband has been concerned with data. Voice connections do not need additional bandwidth, a normal voice channel uses 64 Kbps and provides adequate quality. There is no business pressure to provide anything better.

One use of broadband for voice is to fit more voice channels into a 2 Mbps leased line. A primary rate ISDN line provides 30 voice channels over 2 Mbps. Products are available which use fast packet-switching techniques to increase this capacity further. They send the voice samples as data packets, and transmit packets only when speech information is present. This saves considerable bandwidth since silences in the conversation (for example while listening to the remote speaker) do not use bandwidth. Increases of up to four times are claimed. This seems a worthwhile improvement, but line speeds of at least 2 Mbps must be used to prevent excessive delays to the speech (rather like the effect on a satellite call). It is worthwhile only if between 30 and 120 voice circuits are needed between two sites, which represents a considerable level of traffic!!

5.3 *Specific Techniques*

The earlier part of this paper has exposed a growing need for greater bandwidth. There are a number of products and standards which aim to satisfy this need. There is convergence of the features offered by these, but equally there is competition. There are too many solutions for all to succeed. Inevitably there will be some losers. The situation is similar to the position of LAN standards in the late 70's and 80's. There were then several potential LAN protocols, of which a few (Ethernet, Token Ring) have been very successful, while others (e.g. Token Bus, Cambridge Ring) are not nearly so commonly used.

The following sections look at some of the current broadband technologies. It is difficult to predict which will be commercially successful and which will not since the factors affecting are not purely technical.

The relationship between the various standards is complex and they overlap. Figure 4 shows some of the ways they might interact.

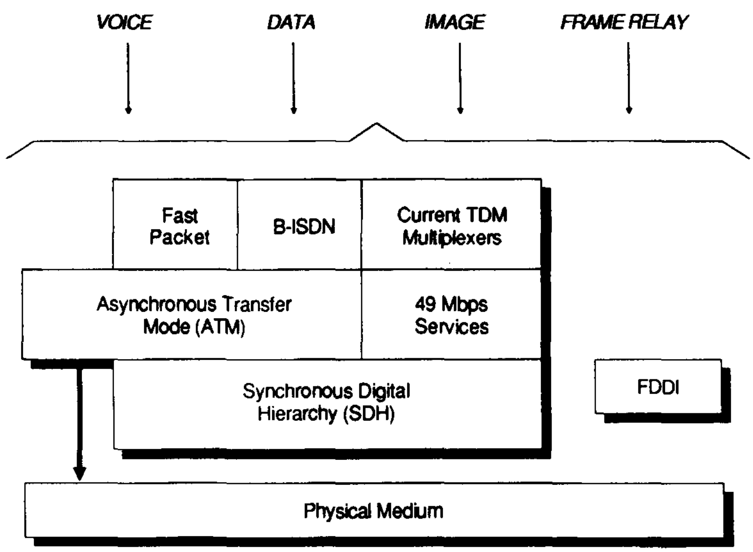


Fig. 4 Multi-services

5.3.1 Frame Relay Frame Relay is an interface standard defined in CCITT recommendation I 122, and covers the transmission of *frames* using ISDN link level protocols. By using link level protocols frame relay avoids the overheads of packet protocols such as X 25. Not all products termed 'frame relay' conform to the CCITT recommendation.

Frame relay is intended as the one of the interfaces which users equipment will use to access other higher speed equipment.

5.3.2 Fast Packet Fast Packet is a technology which has partly evolved from packet switching in that it is a streamlined packet protocol designed to provide switched connection between LANs or to act as a fast 'core' or backbone for a normal packet switch network. Many of the features of packet switching have been removed, including error correcting.

As a result fast packet is capable of very high throughputs, in the order of 1,000,000 packets per second. However since it has no error protection it is best to use it over a reliable transmission medium, and to provide error correction at higher layers in the protocol stack.

Fast Packet switches usually support a number of different input ports including voice, data at 64 Kbps and sometimes 2 Mbps, and frame relay. A fixed packet or 'cell' size is used on the trunks. This is often the ATM cell size (see below) so that fast packet switches can use ATM, although currently most will be connected via 2 Mbps trunks.

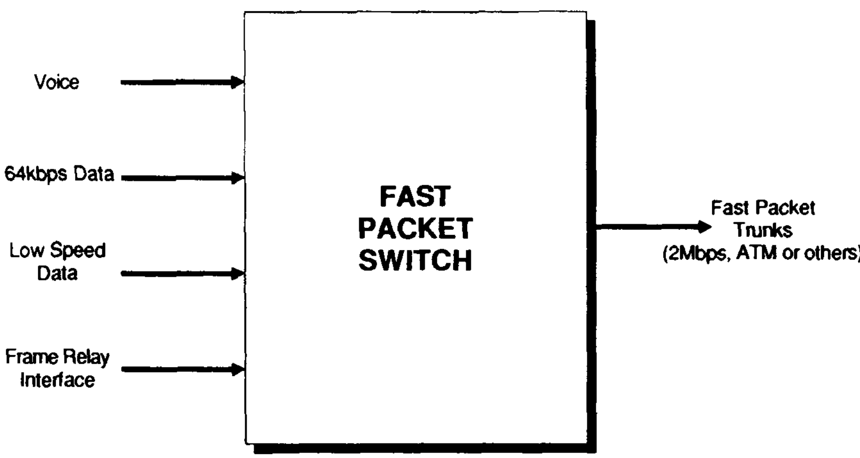


Fig. 5 Fast Packet Switch

5.3.3 Asynchronous Transfer Mode (ATM) ATM is a multiplexing technique using fixed size packets, or 'cells'. A cell consists of a 5 octet header and a 48 octet information field as shown in Figure 6.

The header includes routing information (24 bits), an 8 bit error control field capable of detecting multiple bit errors and of correcting single bit errors in the header (not the information field), and various other fields.

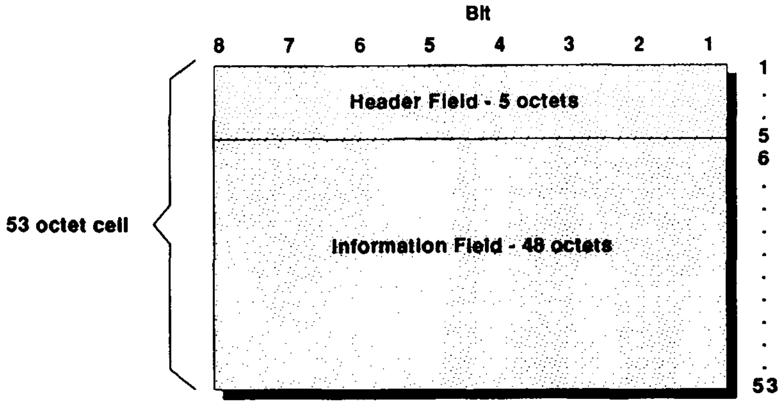


Fig. 6 ATM Cell Structure

ATM can operate directly over a transmission line, but it can also operate at 155 Mbps or 622 Mbps over a physical interface using the Synchronous Digital Hierarchy (SDH). The ATM cell structure is used at lower speeds by some fast packet products.

5.3.4 Synchronous Digital Hierarchy (SDH) The SDH transmission system is defined in CCITT recommendations G707, G708 and G709. It transmits at 155 Mbps, or multiples of that speed, and acts as a flexible, efficient carrier for other services. It can have a number of different frame structures. One is Synchronous Transport Module level 1 (STM1), Whose basic frame structure is a matrix of 270 columns (octets) by 9 rows as shown in Figure 7. This repeats every 125 micro seconds.

This basic frame has control information in the first 9 columns. The remainder of the frame carries 'containers'. These containers are either:

1 × C4 carrying a 150 Mbps service, or
3 × C3 each carrying a 49 Mbps service

The C4 container is also a matrix of 261 columns by 9 rows with the first column for control information. It can carry ATM cells as shown.

5.3.5 Broadband ISDN Integrated Service Digital Network (ISDN) was evolved from normal telephony to provide a high grade digital voice service, plus a 64 Kbps data service. This in itself represented a big increase in bandwidth over that available via modems, but to support the higher bandwidth which users are demanding, ISDN service are moving towards broadband – Broadband ISDN or BISDN.

The proposed BISDN services would make use of ATM transmission to provide on demand bandwidth in excess of 2 Mbps to users. In this sense

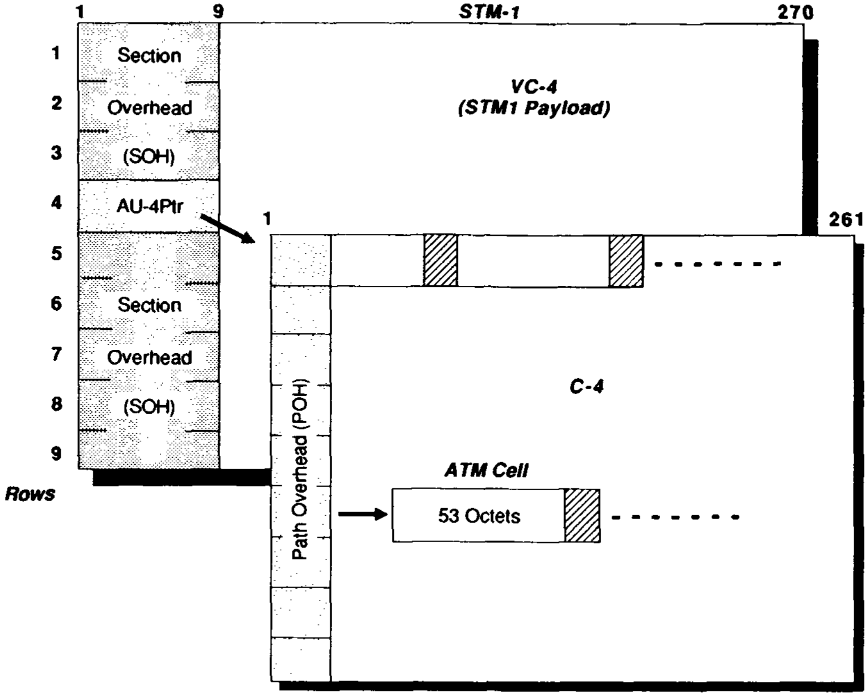


Fig. 7 The STM-1 Frame structure

BISDN is not a technology in itself, but makes use of ATM transmission to provide its service. There are some additional functions to control the BISDN connection and to adapt the ATM interface to the higher layer ISDN functions. These are not shown in Figure 4.

5.3.6 Fibre Distributed Data Interchange (FDDI) FDDI standards provide for 100 Mbps services on a fibre optic LAN and can support voice and image as well as data. These standards are described in more detail in another article in this issue of this journal (Taylor, 1992).

The FDDI-I standard supports data over a fibre LAN covering a range of several kilometres. FDDI-II provides support for voice and other services as well a data. These standards can be classed as broadband since they provide bandwidth for LAN which exceeds that in common use today.

5.3.7 Metropolitan Area Networks Metropolitan area networks (MANs) are capable of supporting a multiplicity of services over a wide geographic area, typically a town as the name implies. The distinction between a LAN and a MAN is becoming increasingly blurred. Proposed LANs using FDDI-II can cover a distance of 100 Km and could equally be considered as a MAN.

Although the technology is converging, there is one important difference between LANs and MANs. LANs are private networks used by a single company to support office and business functions. MANs will be run by PTT, shared by many users and will support a wider range of services-possibly including television and other entertainment services, and telemetry for public utilities.

6 Standards

The International Telegraph and Telephone Consultative Committee (CCITT) is active in defining standards, particularly in the transmission and ISDN areas.

> Recommendations in the G700–G956 cover digital transmission systems and multiplexing equipment.
>
> The I series recommendations cover ISDN topics including Broadband ISDN and Frame Relay.

Another body with an active programme of work in transmission techniques is RACE (Research and Development in Advanced Communications Technologies in Europe). It was established in 1988 by the European Commission with the aim of developing new and innovative telecommunications-based services for Europe. Its programme of work covers a wide range of communications topics; much work on ATM and related techniques has been done as part of the RACE programme.

7 Broadband and the PTTs

The broadband technology described in this paper is applicable equally to private networks or public networks. Approaches such as fast packet or FDDI are targeted perhaps more at private networks, while BISDN and MANs are targeted more at public services. However, the choice of approach will be determined more by the background and organisation of the people working in the fields, than by underlying technical consideration.

Broadband networking, by definition, provides facilities for the movement of large volumes of data. This makes it attractive for large private networks. Smaller users with an occasional demand for high bandwidth may wish to use public services instead. The work on BISDN is being promoted by the PTT's to meet this need.

Much of the work on Synchronous Digital Hierarchy (SDH) is also more applicable to PTT's than private networks. It provides a very high speed service which acts as a flexible carrier for lower speed services. Initially at least these high speed services will be used by the PTT's to provide 2 Mbps services to users, and in the future, new higher speed on-demand services.

Most PTT's and major vendors are actively involved in development of these new broadband services. As yet most services are still in the future, but the demand is beginning to show and service providers and equipment vendors want to be ready.

Reference

FULLER, A.R. "Future Applications of ISDN to Information Technology" *ICL Tech. J.* Vol. 7, no. 3, pp. 501–511, May 1991.

Further Reading

"Open Networks – The Key to Global Success", J. Houldsworth. *ICL Tech. J.* Vol. 8, no. 2, pp. 179–197, 1992.

"Corporate Network Infrastructures Through the Nineties", J. Lawe, A. Flatman, B. Russell. *ICL Tech. J.* Vol. 8, no. 2, pp. 198–209, 1992.

"Open Systems LANs and their Global Interconnection". J. Houldsworth, A. Flatman, M. Taylor, K. Caves, K. Crook. Butterworth Heinemann. 450pp, ISBN 0 7506 1045 X. 1991.

"Realizing Broadband: The Regulatory Hurdles", V. Patel. *Telecommunications* March 1992.

"WAN Internetworking – Emerging Standards and Services", G. Smith. *Inte Net* January 1992.

Biography

John Meers

John Meers is a consultant with ICL Associated Services Division specialising in computer networking. He has over 25 years experience in this field, and his expertise covers data and voice networking and messaging. He graduated from Glasgow University in 1962 and joined STC where he worked on the design of telex and message switching networks. In 1985 he joined ICL where he has been involved in marketing and consultancy relating to computer networking. He has been involved in BSI and OSI committees working on the development of networking standards.

FDDI – The High Speed Network for the Nineties

Mark Taylor
Memory & Technology Applications, ICL Manchester, UK

Abstract

FDDI (Fibre Distributed Data Interface) is an emerging Local Area Network standard with the performance capability to support network users through most of the 1990's. This paper compares the performance of this new generation Local Area Network to current CSMA/CD and Token Ring technologies and extrapolates the trends into the coming decade. It describes the ways in which the FDDI concept has evolved to provide a rich selection of topologies and technologies. These allow the user to tailor a system to their individual needs whilst retaining all the benefits of Open System Interconnection.

1 Introduction

FDDI is a Local Area Networking standard which will be of inestimable importance in the 1990s. It offers a data rate of 100 Mbits/sec, high connectivity, high efficiency and comprehensive network management facilities.

The current generation of Local Area Networks (LANs), while highly successful, is represented by a number of different standard and non-standard approaches. These include CSMA/CD (Carrier Sense, Multiple Access with Collision Detection), Token Ring, Token Bus and Slotted Ring standards, with non-standardised contenders such as Arcnet also having significant market share. This multiplicity has the effect of weakening the effectiveness of individual standards, confusing users and creating problems with interworking. FDDI on the other hand is the only LAN in its class of performance which is currently being developed as an international standard. Although this in itself produces an additional LAN type which must still interwork with those of the current generation, it at least gives confidence that further proliferation of standards may not occur. Also the interworking problems are now well understood and are catered for. The current generation of LANs are not easily extensible to the data rates provided by FDDI and thus a new development was unavoidable.

2 Local Area Networking Trends

Just as in every other area of Information Technology, the trend in Local Area Networks is toward increased functionality and speed of operation. This is hardly surprising given the exponential growth in power of the end systems which they support. The projected performance trend for single chip micro-processors against time of introduction, indicates that their performance will be greater than 2000 MIPS by the year 2000 compared to around 20 MIPS in 1990. Memory capacity also can be shown to grow at an exponential rate when plotted against time.

Clearly the applications spawned by these staggering increases in end system performance will put pressure on the LAN interconnections to provide similar performance improvements. Coupled with this are increasing expectations for system reliability, resilience and manageability.

LANs, although with fewer available data points to plot, appear to echo these trends in terms of increased data rates. An approximate curve is shown in Figure 1.

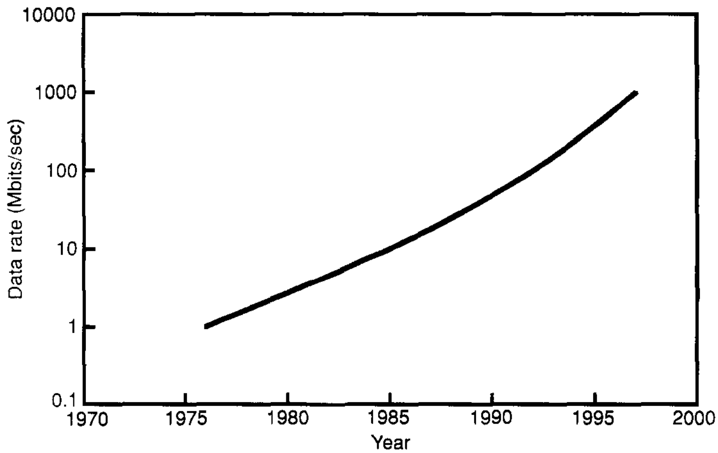


Fig. 1 LAN Data Rates versus Time

LANs only began to emerge from research establishments in the late 1970's at which time 1 Mbit/sec was a typical data rate. Heavy deployment of Ethernet at 10 Mbits/sec and Token Ring at 4 Mbits/sec began in the mid 1980's whilst FDDI at 100 Mbits/sec is following a similar pattern in the 1990s but possibly a couple of years earlier in the decade. FDDI development was in fact started in the early 1980's but until recently the areas of applicability, costs and state of development have limited its uptake.

The significant data points in Figure 1 are at 1985 when the basic set of CSMA/CD LAN standards at 10 Mbits/sec were approved. At this time working chip sets were available and the deployment of the LAN took off. Similarly in 1992 the final part (Station Management) of the basic set of FDDI standards at 100 Mbits/sec should be approved and the knee in the curve of product shipments should be reached. Looking into the future, FDDI Follow On LAN (FFOL), which is in the study phase, is optimistically predicted to be standardised at around 1995. This LAN will probably operate at 622 Mbits/sec.

The important role of standards in the LAN arena limits the number of data points available for plotting this type of graph. The data rate increases in order of magnitude steps. Had there been a market free-for-all, a relatively smooth line of points would have been available to confirm the trend, but on the other hand the market would probably not be so well developed.

Despite these steps in capability of the service provided, the deployment at a particular step has been evolutionary rather than revolutionary. A progressive broadening in the areas of applicability has occurred as costs have fallen and operational experience has increased. Thus CSMA/CD ports appeared initially on mainframes, minis and high performance workstations before starting to be introduced on top of the range PCs and before finally becoming cost effective on even the lowest end systems. This same trend is clearly observable for FDDI connections, except that FDDI supports an additional application area, that of a backbone to existing LAN types. These advances into more widespread application areas have been encouraged by extensions to the standards from the initial media type to lower cost alternatives.

Over utilisation of large LANs can be countered initially by splitting them up into a number of smaller LANs connected by MAC bridges. There is a relationship between distance and communication bandwidth requirements. Network applications tend to be shared between local work groups with communications to more remote parts of a site being less frequent. MAC bridges can prevent traffic which does not need to leave a particular LAN from being passed to another. In this way a bridged network can support far more traffic in total than would be possible on a single large LAN. The normal trend in a large networked site is for a single LAN to grow in size to a point where it can no longer carry all the traffic. It will then be subdivided into smaller localized LANs based around work groups. These will be bridged to a backbone LAN which links them all together (The bridges will often replace repeaters in the original LAN.) Eventually as the number of bridged work group LANs grows and the traffic on these increases, it becomes necessary to replace the backbone LAN with something faster. This is where FDDI will initially come into play in many organisations. In time as FDDI is increasingly deployed at the work group level, the FDDI backbone will itself become overloaded and will probably be replaced by FFOL towards the end of the decade.

Some of the more enthusiastic FDDI advocates have predicted that FDDI will sweep away CSMA/CD and Token Ring. This however is analogous to CSMA/CD replacing the ubiquitous RS232 port. The latter is still prevalent on the back of most items of IT equipment (though given a couple of years it is quite possible that 10BaseT, the twisted pair implementation of CSMA/CD could begin to replace RS232 as a printer interface). A hierarchy of LANs at different performance levels will exist in most organisations which are suited to the applications which they support. FDDI's time has come however and network users not planning for its introduction could well find themselves struggling with overloaded LANs.

3 Comparison Between FDDI and Other LANs

So what are the differences between FDDI and the prevalent LANs of the moment, Ethernet and Token Ring?

Table 1 Comparison of different types of LAN.

PARAMETER	FDDI	CSMA/CD	TOKEN RING
Data Rate	100 Mbits/sec	10 Mbits/sec	4 or 16 Mbits/sec
Baud Rate	125 MBaud	20 MBaud	8 or 32 Mbaud
Max Frame size	4500 Bytes	1500 Bytes	4399 Bytes
Max Number of Stations	500 to 1000[1]	1024	72 to 250[2]
Physical Topologies	Ring, Star	Bus, Star	Ring, Star
Logical Topology	Ring	Bus	Ring
Max pt. to pt. length, i.e. station to star or station to station without repeaters	2 km – Fibre 60 km – SM fibre 100 m – STP* 100 m – DG UTP*	1 km – Fibre 600 m – Coax 185 m – Thin coax 100 m – UTP 2km – Fibre*	2 km – Fibre* 75→173 m – STP (16Mb/s) 100 m – DG UTP (16Mb/s)*
Max station to station separation	50 km (Flattened ring of 100 km circumference)	3 km approx	<1 km (Star topology)

[1] 1000 physical links which may correspond to 500 to 1000 stations dependent upon type.
[2] 72 for telephone twisted pair cables, 250 for IBM type 1 cable.
*Not yet an approved part of the standard.
DG UTP = Data grade unscreened twisted pair.
STP = Screened twisted pair (IBM Type 1).

Table 1 compares some of the basic parameters of these three LAN types and is followed by a discussion of some of these together with other important features which are less easily summarised.

From the above table it is obvious that FDDI provides the connectivity of CSMA/CD LANs without the distance limitations. This it does at a raw bit rate which is an order of magnitude greater.

3.1 Data Rate Comparison

An increase in the basic bit rate for a newer LAN type is one improvement which is to be expected as the speed capabilities and integration density of electronic components increases with time. The decision to use optical fibre as the transmission medium also allowed increases in speed due to its greater bandwidth/distance product relative to copper cables.

3.2 Connectivity Comparison

The bus structure of CSMA/CD provides this interface with high connectivity. Token Ring on the other hand could never support a similar number of stations on one ring. The main reason for this is accumulated phase shifts around the ring which are also known as jitter. On a token ring there is one selected station (the monitor station) which is responsible for providing the clock for the entire ring. All other stations employ a phase locked loop to recover this clock from the received symbol stream and then use it to retransmit. Low frequency components in a data stream which can be tracked by the phase locked loops, together with other sources of drift, conspire to modify continuously the length of the ring in terms of the number of bits which it can hold. As this would otherwise cause bits to be suddenly added to or subtracted from circulating frames, the monitor station employs an elastic latency buffer to provide compensation. By continuously adjusting its length, the buffer maintains the ring at a constant length. The depth of this buffer therefore determines how many stations can exist on a ring. Specification inadequacies in the standard have prevented even the specified number of stations from being deployed on Token Rings and this problem has yet to be resolved.

FDDI on the other hand circumvents these problems by reclocking the data in each station. An elastic buffer of limited length is still needed in each station simply to accommodate the slight differences in clock frequency between stations. The elasticity of this buffer imposes a limit on frame lengths (approximately 4500 Bytes) but not on the number of stations.

3.3 Topology Comparison

Collision detection requirements of CSMA/CD LANs limit the point to point distances which are possible. A minimum frame length ensures that if a collision occurs between two stations at far extremes of the LAN, then the overlap will always be detectable by any involved stations. This overlap fails to occur if the network size is over extended. A minimum frame length can be wasteful of bandwidth if there is insufficient data to fill it. Thus a compromise must be reached between network size and efficiency. On the other hand the star and bus topology allows minimal latency between any two points on the LAN (in the absence of contention) by avoiding the requirement for a frame to visit every station in turn as it does on a ring.

Token Ring, although logically a ring, is rarely configured this way physically. This is because a break in a single ring puts all the stations out of action. Although a relay at the junction between the ring and the branch (lobe) cables is used to by-pass broken or powered down stations, it does not protect against cable breaks. To combat this problem, the relays (Trunk Coupling Units) are grouped into concentrators so that the ring 'medium' is reduced to connections within a box and the branch cables hence form a star configuration. For the existing standard, the concentrator is a passive device; for future enhancements to the standard these will be active devices.

FDDI provides two solutions to the broken ring problem. The first of these is to run a redundant ring in parallel with the one in normal use. The direction of flow in this redundant ring is opposite to that in the primary ring. This configuration, known as a reconfiguring dual ring, allows stations at either side of a problem to form a connection between the primary ring and the redundant secondary ring under the control of Station Management. In this way an elongated 'ring' is established with the stations to either side of the break at extreme points logically when they were previously adjacent. The second optional strategy protects against broken or powered down stations by means of an optical by-pass relay which is an optical analogue of an electrical relay. Note: Token Ring standards have now adopted the reconfiguring dual ring strategy in an optional extension.

As was the case for Token Ring, there is also the option of a star configuration using a concentrator device. For FDDI, concentrators are always active devices which provide electrical by-passing. The signal power losses of optical by-pass relays are too great to allow a passive analogue of the Token Ring concentrator. Concentrators can be cascaded in tree configurations and can also form a part of a reconfiguring dual ring. Note however that the tokens and frames still pass from station to station in a sequence determined by a logical ring which is mapped across a physical star topology. Figure 2 below demonstrates this arrangement.

FDDI therefore provides a flexible choice of physical topologies which can include both rings and stars in whatever combination best suits the individual site. This choice can be made without sacrificing the resilience of the ring.

3.4 Latency Comparisons

Latency is a parameter which was not included in the list of Table 1 as there is no simple value which can be considered. It is an important aspect of any LAN Medium Access Control (MAC) method however and one which can be a disadvantage for FDDI unless care is taken to use the LAN in an appropriate manner. This discussion does not take account of the queuing

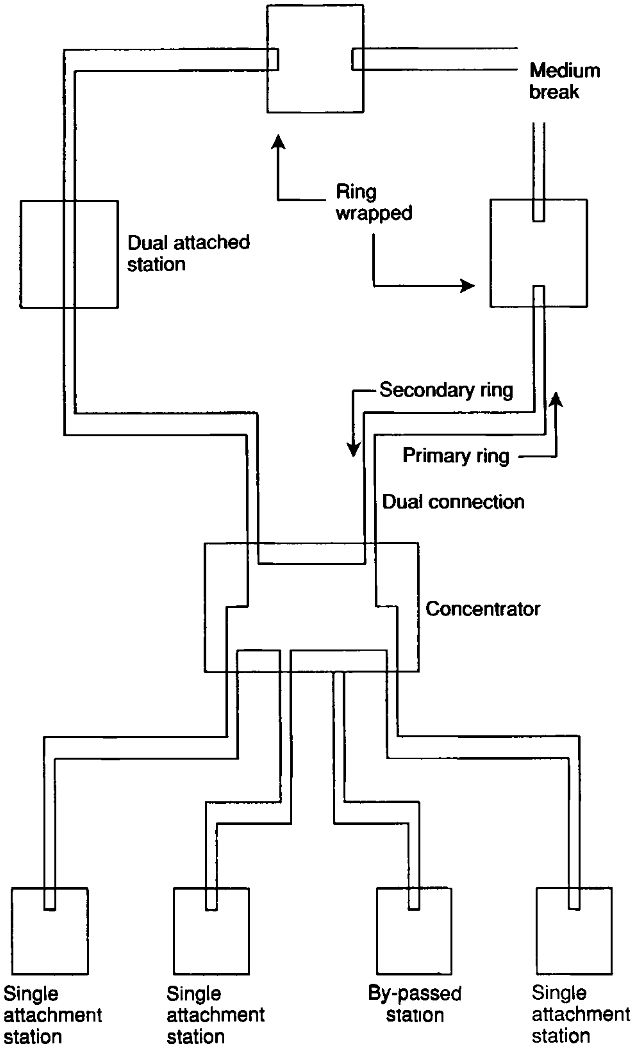


Fig. 2 Illustrating the mapping of a logical ring onto a physical star and ring topology

delays which occur on any network where frames are generated in an unschedualed manner. A frame (or group of frames where they are grouped for transmission as may be the case for FDDI) is assumed to arrive at the MAC function as soon as generated.

CSMA/CD LAN stations do not schedule their accesses to the media; they simply transmit when there is no activity detected on the medium. Provided that a collision does not occur, the frame travels directly to the receiving station (and all other stations) via a small number of repeaters which add little delay. Latency is therefore low except when collisions start to occur and stations then have to back off for a period. This backing off becomes more frequent and longer lasting as the total network activity increases. It is necessary to limit the traffic on a LAN to less than it can theoretically sustain if the low latency characteristics are to be retained.

FDDI passes a token from one station to another and the holding of this confers the right to transmit. Each station sequentially has the opportunity transmit for a period of time which is related to the stations assessment of the level of activity on the LAN. As each token or data frame passes sequentially from station to station, intrinsic delays are generated at each of these and also by the paths along the media, but most significantly by the time for which the stations along the ring transmit their data.

The time for which a station can hold onto the token is controlled by a Timed Token Protocol [Grow 1982]. This creates a time limit for the token to complete a full revolution of the ring. Each station measures the time from which it last received the token and uses the difference between this and a predetermined time value (TTRT, the Target Token Rotation Time) to determine for how long it may transmit. The TTRT value is the same for all stations. If a station uses all the available time then the subsequent stations are left with no transmission time and the token is passed quickly from station to station. The Token then completes the circuit in minimum time to the station following that which last transmitted. This station, having started its token rotation timer after the end of the previous stations long transmission, now has most of the TTRT time available in which it can transmit. When a station does not use its full time allocation then subsequent stations have the remainder in which to transmit without having to wait for the token to complete a full circuit.

The above briefly describes the protocol for asynchronous bandwidth allocation. An additional optional protocol known as synchronous bandwidth allocation can overlay this. This allows stations a limited time to hold the token when it is passed to them regardless of whether time is available to transmit in the asynchronous mode. The synchronous bandwidth is shared out between stations such that if they all make full use of it, then the token will take a maximum of an additional TTRT time to circulate the ring. Synchronous bandwidth is an optional part of the standard and is not covered by International Standard Profiles (ISPs). It also needs to be managed on a per ring basis and is therefore not a plug and play process. These factors may constrain the majority of rings to the use of asynchronous protocol.

The asynchronous transmission protocol may result in a time of approximately N * TTRT between accesses to the medium for a particular station if N stations are making full use of their available transmission time. TTRT has a large range of possible values. It is set by a bidding process between stations at initialization time in which the station bidding the lowest value wins the right to initialise the ring with its value. It is preferable for a low value to be selected in order to minimise latency. If too low however, the efficiency of the LAN will be reduced dramatically. In the limit, if TTRT equals the time for the token to rotate in the absence of data transmissions then the token will spend all its time going around the ring without any station being able to use it. A lower limit is therefore set at 4 ms which avoids this conundrum. In practice the value often recommended and used is 8 ms [Jain 1990] which gives a minimum of 90% efficiency on a large ring. With 500 stations on a ring and this value of TTRT, a station may have to wait as long as 4 seconds between accesses to the medium. When it does gain access however, it has the opportunity to transmit almost 100 kbytes i.e., more than 20 maximum length FDDI frames or more than 60 maximum CSMA/CD frames.

Clearly a different access philosophy may be needed from that for CSMA/CD, where frames can be forwarded to the LAN as they are generated. FDDI stations may have to store a large number of frames between accesses to the LAN.

The performance achieved by FDDI stations can depend on the implementation of the transport layer protocols and of their frame buffering capabilities. If it is necessary to transmit a single frame and then wait for an acknowledgement, then a station will only be able to transmit a single frame whilst it holds the token. A similar state of affairs will exist if the station has only limited frame buffering capability. If on the other hand the transport implementation is efficient and can bundle a large number of frames per acknowledgement and if the frame buffers are capable of holding all these, then maximum use may be made of each access to the medium.

If some stations on a particular LAN are capable of making full use of the token holding time, then those which can't may be found to achieve a very low throughput. Take the example of a station which is attempting to transmit 1 kbyte frames and requires an acknowledgement for every frame. On the same ring there are 100 stations which hold on to the token for the full TTRT time of 8 ms. The station will achieve a throughput of only 1·25 kbytes/sec (one access per 0·8 seconds)! There are two situations which could make this throughput even worse. The first occurs if the receiving station is passed the token immediately after the data and cannot turn around the acknowledgement quickly enough to transmit it before passing on the token. In this case the throughput will be halved. The second is when the network is extended by MAC relay bridges and the receiving station is on a different sub-network. In this case each bridge must wait for access to the LAN to which it requires to transmit. Each frame and acknowledgement suffer the delays of all the sub-networks through which they have to pass.

On a CSMA/CD subnet (a single LAN which is not bridged) the latency delays plus acknowledgement duration (ignoring the station response time) only total about one eighth of the transmission time of a 1 kbyte frame.

Furthermore each individual frame has to contend for access to the medium. There might thus be little difference in performance between an efficient and inefficient implementation of transport protocol.

3.5 Efficiency Comparisons

Efficiency of the MAC protocol as opposed to that of the higher layers, which are independent of the LAN type (but may interact with it), was touched upon briefly in the previous section. The efficiency is the fraction of the user data throughput of the entire LAN divided by the raw data rate of the LAN (100 Mbits/sec for FDDI). As each of the higher layers may add its own overheads in terms of frame headers, acknowledgements etc, in the context of the MAC protocols, these overheads are taken to be part of the user data. As the MAC frame headers and enforced interframe gap times are of a similar number of bits in all the standard LANs, these can also be ignored for the purposes of a comparison.

The major factor controlling the efficiency of an FDDI LAN is the ratio of the Target Token Rotation Time to the actual time which the bits take to travel around the ring (D). For small LANs or for large values of TTRT the

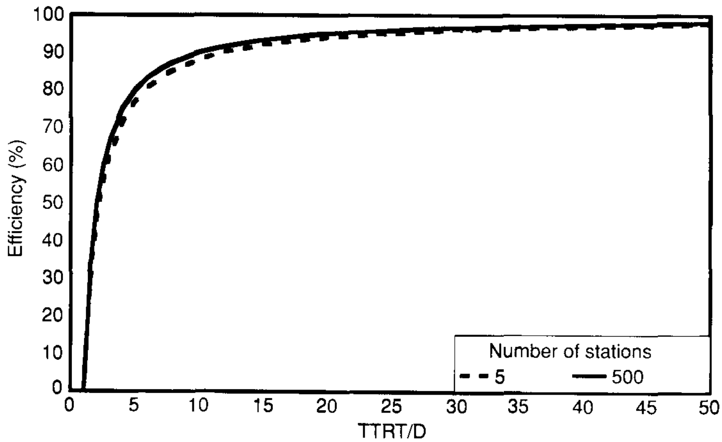


Fig. 3 FDDI Efficiency versus ratio of TTRT to D.

efficiency approaches one as shown on Figure 3. There are other overheads associated with the MAC protocol. Irregular events such as stations joining or leaving the ring cause the ring to reinitialise and this is quite a time consuming process. Some of the overheads of station management are strongly associated with the MAC protocol i.e., they might not be necessary in the management of another type of LAN. One example would be the Neighbour Identification Frames which each station transmits at regular intervals to determine its nearest Neighbour. In general however these overheads will have a negligible impact on the overall efficiency of a correctly operating LAN.

Token Ring MAC protocols are very similar to FDDI. The timed token rotation protocol is not used however. The stations simply have a set token holding time. In itself this is very efficient, the only 'dead-time' being that of the token itself which is very short compared to data frames. One major cause of inefficiency does exist however. A station cannot release the token until it has received the start of the last frame which it transmitted. If the ring is large and the frames short, this results in a significant amount of dead-time on the ring. An early token release option has been added to the standard to get around this problem but a price to pay is that it prevents the priority setting mechanism from working properly.

CSMA/CD efficiency is very dependent on the network size. Dead-times occur during collisions. These result from the possibility that Station A might start to transmit when Station B has already started but the transmission from this has not progressed far enough through the network to be detected by Station A. Collisions may also occur when two or more stations defer to some coincident network activity then transmit simultaneously when this activity ceases. (This is the behaviour in protocols known as a 1-persistent which include CSMA/CD.) In the first type of collision, if the network delays can be reduced to zero, then there will be no collisions due to this cause. In the second type of collision, the collision is detected quickly if the delays are short and the retransmission attempts can quickly be reschedualed. The efficiency with short network delays and long packets can therefore exceed 0·9. On the other hand with maximum network sizes and minimum packet lengths the bandwidth utilization can fall to below 0·5. In theory the efficiency may be as low as 0·37 but this assumes minimum sized packets and an impossible topology where every station is separated from every other by the maximum delay. In practical networks the efficiency will be in the range 0·5 to 0·8. Some additional loss in efficiency results from the minimum frame size restriction (64 bytes including MAC overheads).

3.6 Deterministic Access

Despite the complications concerning throughput, FDDI guarantees fair access at the MAC layer (this can be overruled by priority settings) and access within a bounded time interval. This deterministic access as it is termed, is not possible with CSMA/CD LANs. The randomness of access attempts and of the collision recovery process, mean that a statistical possibility exists that a station may never succeed in accessing the medium or that the time to do so may exceed acceptable bounds. In some process control applications deterministic medium access is considered essential and FDDI thus satisfies these requirements.

3.7 *Management Comparisons*

The management facilities of FDDI, defined by the Station Management (SMT) draft standard, are the most comprehensive of any LAN so far. At the time of writing, draft 7·1 is available and is not expected to undergo any significant changes before being approved. This represents the last part of the basic set of FDDI standards to be completed and runs to a daunting 200 plus pages in several major sections.

Connection Management is subdivided into the sections, Physical Connection Management, Configuration Management and Entity Coordination Management.

Physical Connection Management controls the point to point links which together form the ring. Determination of a good or an absent connection allows station management to change the configuration to avoid a bad link or make use of a good link. Information regarding the state of links can be conveyed to the managing entity so that remedial work can be initiated.

Configuration management controls the interconnection of PHY (Physical Layer) entities and MAC entities within Stations and concentrators such that appropriate by-passing and ring wrapping actions can be achieved and a continuous logical ring maintained. The configurations are made by the Configuration Control Element (CCE) which is the path switching function within a station. The nature of the CCE depends on the type of station. In a concentrator for instance it will need to have the capability to by-pass ports which are giving problems. In dual attachment stations it will need to provide the ring wrapping function at each of the two (A & B) ports to bring in the redundant counter-rotating ring. Alternatively a Dual Attachment Station can be connected to two ports of a concentrator with one of the connections withheld and used as a standby for resilient operation. More sophisticated configuration capabilities may be included in some implementations such as the ability to attach additional MACs. Provided that care is taken to avoid undesirable connections a wealth of different configuration possibilities exist under the control of network management.

Entity Coordination Management controls the operation of the optical bypass relay if present and coordinates this with the Physical Connection Management. It also carries out some diagnostic functions.

Ring management receives and reports on MAC specific aspects of the ring status and takes actions as a result where necessary.

Frame services provides peer to peer frame formats and protocols to allow cooperative behaviour and information passing between different SMT entities on a ring and upward communication with a management agent.

SMT has been formatted in a manner which follows the approaches defined by standards for layer management, unlike earlier LAN management standards which were written in a more ad-hoc manner and which in many cases are now having to be re-written for acceptance as international standards.

Management for CSMA/CD has been largely an afterthought. The first issue of the standard had a blank section for layer management. This has largely now been rectified with standards for Layer (i.e. Station) Management and repeater management, although these will not generally be easily applied to the large installed base of equipment.

Token Ring has always been considered to have good management facilities. The Token Ring access method would be difficult to operate without at least some basic management because of the need to recover the ring from lost tokens, to introduce and remove stations and to locate problem points. FDDI Station Management was developed on similar lines to Token Ring Management and it shares a number of the same facilities. It provides far more functionality however particularly in terms of configuration management and is specified in a way which allows easier integration into standard management systems.

4 FDDI Technology Progression

Apart from the concentration on a particular Medium Access Control method which FDDI should provide in the present decade, there is also a trend towards a unified approach to network cabling. Cabling standards are currently being written to define a limited selection of media types and of installation topologies for these.

Star configurations of a hierarchical nature connect wiring centres together and provide the final link down to the user access point. This is illustrated in Figure 4 below.

FDDI concentrators fit well into such topologies and can be housed at the wiring centres.

For the intermediate wiring centres the multimode optical fibre of the existing FDDI standard is highly appropriate as the distances may be great and noise environment unfavourable. At the last stage to the desk however, there has been little enthusiasm so far for the installation of optical fibre and the available medium is generally twisted pair cable. The reason for this reluctance is the cost of such installations when set against any clearly perceived short term needs for the widespread provision of 100 Mbits/sec to the desk.

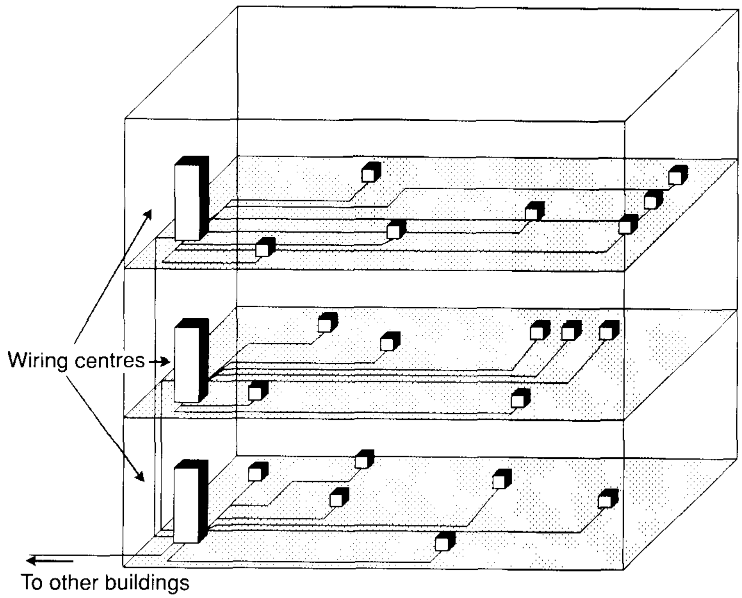


Fig. 4 Structured cabling example.

4.1 FDDI on Twisted Pair Cables

As a cost reduction exercise and to address the existing installations of twisted pair cables, an initiative has been started to map FDDI (referred to in this case as TPDDI) onto this medium. This effort is still in the early stages and is complicated by the existence of three different types of twisted pair cables. These are old style Unscreened Twisted Pair (UTP) telephony type cables, high performance 'Data grade' UTP cables and Screened IBM Type 1 cables. At the time of writing this article, it has yet to be proved conclusively that Electromagnetic Compatibility (EMC) requirements can be adequately met.

Even if this does not prove possible there is another way in which a good substitute may be provided over twisted pairs to the desk.

A number of vendors are offering versatile hub devices for use at wiring centres. These house a variety of plugable LAN cards so that any LAN type concentrator or multiport repeater or a mixture of these can exist in the same unit. The continued integration of the MAC bridging function has also allowed multiport MAC bridges to be implemented on a card so that dissimilar LAN types can also be interconnected within a hub. By multiport bridging of an FDDI LAN to multiple CSMA/CD 10BaseT LANs, the hubs can provide stations with individual twisted pair LANs at the full CSMA/CD bandwidth of 10 Mbits/sec. 10BaseT has proven EMC performance on twisted pair cables. As any stations which individually need more than 10

Mbits/sec of bandwidth will soon overload an FDDI LAN, unless used in very limited numbers, this pragmatic solution will suit most applications should the TPDDI initiative flounder.

4.2 Low Cost Fibre FDDI

A lower cost version of the fibre standard is also being developed with the intention of encouraging more fibre to be used within structured cabling installations and of encouraging the use of FDDI LANs where the fibre has been installed. This version has a reduced power budget, a limited distance span and a different, smaller connector. The connector will enable cost reductions as a consequence of both lower part costs and higher integration densities and the reduced power budget will reduce the cost of the active optical components and their housings.

4.3 Single Mode Fibre FDDI

There are instances where it is required to extend an FDDI LAN beyond the confines of a single campus or where the distances between stations on a site exceed 2 km and it is not feasible to add a repeater.

To address these situations, an extension of the standard to cover single mode fibre operation has been added.

Single mode fibre permits an order of magnitude increase in transmission distances relative to multimode fibre. Such distances make it possible to use FDDI as a Metropolitan Area Network (MAN) and this was a main incentive for producing the Single Mode option. Another important application is that of disaster stand-by systems. In these, all the data on a system is duplicated at a remote site so that catastrophic failure at the primary data processing site does not result in a total loss of operational data.

4.4 FDDI Follow On LAN (FFOL) III

Before one standard is complete it is necessary to start looking to what follows. The gestation period of standards can be considerable. CSMA/CD took 12 years from the initial Ethernet concept to an approved standard (1973 to 1985). FDDI has taken 10 years (1982 to 1992). As indicated by Figure 1 the time for the next generation, represented by FFOL, needs to be halved if the target date of 1995 is to be achieved (FFOL was started in 1990). This will present a considerable challenge as the complexity of this standard could potentially be far greater than anything which has gone before. One of the challenges may be to resist the pressures to add features and options which are not essential.

The aim for FFOL is to provide a network service for both the traditional computer generated data, multimedia applications and also for traffic emanating from the Wide Area Networks of the future, the Broadband ISDN networks using Asynchronous Transfer Mode (ATM) cell-based switching architectures. The need to interface effectively with SONET (Synchronous Optical NETwork) public network carrier services, has dictated a transmission rate of 622 Mbits/sec [Ross/Fink, 1992].

This will require interfaces to both isochronous services and asynchronous services via isochronous and asychronous MAC functions and a means to multiplex between the two. Some experience has been gained in this area with FDDI II which should help the process.

No firm approaches have yet been decided on Medium Access Control methods. That which has received most exposure is a slotted register insertion ring. This is being vigorously promoted by IBM. Register insertion rings can achieve a high throughput because of the in-ring storage. The latency however, can be high and this may not suit some potential LAN applications, for example those running coherency protocols.

Summary

FDDI is the focus of Local Area Networking for the 1990s. It provides a number of advantages over existing LAN types which are summarised in this paper. These include higher bandwidth, connectivity and distance together with manageability and a wide choice of physical topologies and media types. The standards for FDDI are near completion and it is approaching a point where large scale deployment will occur. It is no longer an interesting concept for the future, but should be a part of the plans of most network managers.

Acknowledgements

ARCNET is a trademark of Standard Microsystems Inc.

ETHERNET is a registered trademark of Xerox Corporation

IBM is a registered trademark of International Business Machines Corporation

References

GROW, R.M. "A Timed-Token Protocol for Local Area Networks," Presented at Electro '82, *Token Access Protocols*, Paper 17/3, May 1982.

JAIN, R. "Performance Analysis of FDDI Token Ring Networks: Effect of Parameters and Guide-lines for Setting TTRT", *Proc. ACM SIGCOMM*, pp. 264–275, 1990.

ROSS, F.E. and FINK, R.L. "Overview of FFOL – FDDI Follow-On LAN", *Computer Communications*, Vol. 15, no. 1, Jan/Feb 1992.

Standards References

ISO/IEC 8802–3 Carrier sense multiple access with collision detection (CSMA/CD)
ISO/IEC 8802–4 Token passing bus
ISO/IEC 8802–5 Token ring
ISO/IEC 8802–7 Slotted ring
ISO/IEC 9314–1 Fibre Distributed Data Interface – Token Ring Physical Layer Protocol (PHY)
ISO/IEC 9314–2 Fibre Distributed Data Interface – Token Ring Media Access Control (MAC)
ISO/IEC 9314–3 Fibre Distributed Data Interface – Token Ring Physical Medium Dependent (PMD)
IEEE Std 802.3i-1990 10Base-T (supplement to ISO/IEC 8802–3:1990)
pDISP ISO/IEC 10608–6 Definition of profile TA54 for operation over an FDDI subnetwork
pDISP ISO/IEC 10608–14 MAC, PHY and PMD Sublayer dependent and Station Management requirements over an FDDI LAN subnetwork
ANSI/EIA/TIA-568–1991 Commercial Building Telecommunications Wiring Standard

Biography

Mark Taylor

Mark Taylor gained a College Associateship in Electronic Engineering from Bolton Institute of Technology in 1975. He joined ICL in 1977 and has been involved in Local Area Network design and specification for the past 10 years. Since 1986 he has been an active participant in LAN standards committees, chiefly in the IEEE 802.3 group defining CSMA/CD standards until recently when he has represented ICL at the ANSI X3T9.5 committee for FDDI development. He has been involved in the advanced optics part of the Esprit research program and is a member of the team developing ICL's high performance Macrolan fibre optic network. He has co-authored a recently published book on Open Systems LANs.

The Evolution of Wireless Networks

A. V. Flatman[1], **and P. G. Weiser**[2]
[1]OPEN*framework* Division, ICL, Kidsgrove, UK
[2]Logistics Division, ICL Stevenage, UK

Abstract

Wireless Networks are emerging as a significant open standards communications technology, both locally within the building or campus and also within the wide-area telecommunications infrastructure. They will have a number of roles to play within the corporate network, as an alternative to cabled networking, as a means of supporting IT applications in situations where cabling is not practical, or as a supplement to cabling to meet particular networking requirements.

This paper reviews developments in Wireless Network technologies, standards and regulations, and considers a broad range of potential applications together with associated business benefits.

1 Trends

During the 1980s we witnessed the dramatic take up of mobile telephony, with a number of spectacular successes and failures in the public telecommunication services arena. The undoubted successes are the national cellular voice services being operated worldwide; success despite the generally poor quality of service. The most notable failure so far has been the "Telepoint" service, a lower-cost system aimed at the pedestrian and providing coverage within popular public zones such as shopping malls, concourses, etc for outgoing calls only. Opinion points to limited functionality and availability plus overlap with existing cellular and phone box services as the reasons for failure.

Data services are increasingly supported via existing voice-oriented cellular networks, in particular facsimile. Two-way messaging services are also emerging within dedicated mobile data networks. Data will remain the poor relation until high speed, digital infrastructures are established later in the 1990s.

The 1990s are already seeing the deployment of Wireless Networks within the building, albeit limited. Local Wireless networks are being used in environments where cable is either impossible to install, prohibitively expensive or disruptive. Wireless LANs are now beginning to emerge to provide users with a high bandwidth, high integrity data service.

These trends reflect a need to support the increasing community of mobile users within our service-oriented economy; users who need to continue to exploit Information Technology and Telecommunications for business benefit. There is also an inevitable trend towards personal communication and converged communication services (voice, data, image, video), but the rate of implementation is presently unclear.

2 Technologies

Due to distance capabilities and ease of use, radio communications is the most appropriate medium for all public Wireless Network services and the majority of cross-campus and within-building systems. Infrared light may also be used to form very local point-to-point links and networks, limited in practice to line-of-sight. The major advantage of infrared is that it does not require a license. Radio, on the other hand, is required to operate within a strict regulatory regime (more on this in Section 3 below).

It is absolutely essential that precious spectrum (or bandwidth) is utilised in the most efficient way possible. It is therefore worth reviewing the relative merits of the various transmission technologies and modulation/encoding schemes which are appropriate for local and wide-area networks.

Choice of radio frequency band, modulation technique and channel access scheme are fundamental considerations.

2.1 Radio Frequency Band

Frequency bands ranging from several hundred MHz to 60 GHz are allocated across the world for voice and data services. Many of these bands are licensed, with strict requirements imposed for operation in both licensed and unlicensed bands. Frequency management is addressed in Section 3 below however it is worth highlighting here that higher frequencies are less able to penetrate dense materials used in buildings. The latter will significantly affect the choice of frequency band for in-building use.

2.2 Modulation

The radio design engineer will select from Amplitude Modulation (AM), Frequency Modulation (FM) or Phase Modulation (PM) when coupling the Information Channel to the Radio Carrier; each will provide a certain level of efficiency and integrity at the cost of complexity. Efficiency is sometimes referred to as "bandwidth occupancy", or the number of bits/sec supported per Hz. Signal integrity refers to the bit-error-rate of digital signals or the signal-to-noise ratio of analogue signals. Frequency Shift Keying (FSK) and Phase Shift Keying (PSK) are the main categories of digital modulation, but there are many variations; for example instead of having 2 states (i.e. binary shift keying), 4 or more phase positions or frequency shifts are used to encode 2 or more data bits. These variations enable different levels of efficiency and integrity. Direct AM modulation will generally be used for Infra-red transmission for reasons of cost and simplicity.

2.3 Channel Access

It is possible to organise access to a wireless channel in different ways depending upon the application. The two most common schemes are Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA). FDMA will subdivide the total available bandwidth into a number of traffic channels. FDMA is an ideal access mechanism for voice traffic due to the fixed circuits being provided. TDMA divides the channel into time slots, which may be combined to give higher data rates. TDMA is often used in conjunction with FDMA to further increase the granularity of the available bandwidth. Packet-switched data is supported more efficiently via TDMA due to its bursty nature.

Recently advances have been made in the field of Code Division Multiple Access (CDMA). In this scheme the total available bandwidth is allocated to a single wideband channel. Multiple access is accomplished by mixing individual signals with a Pseudo-Random Bit Sequence (PRBS) as illustrated

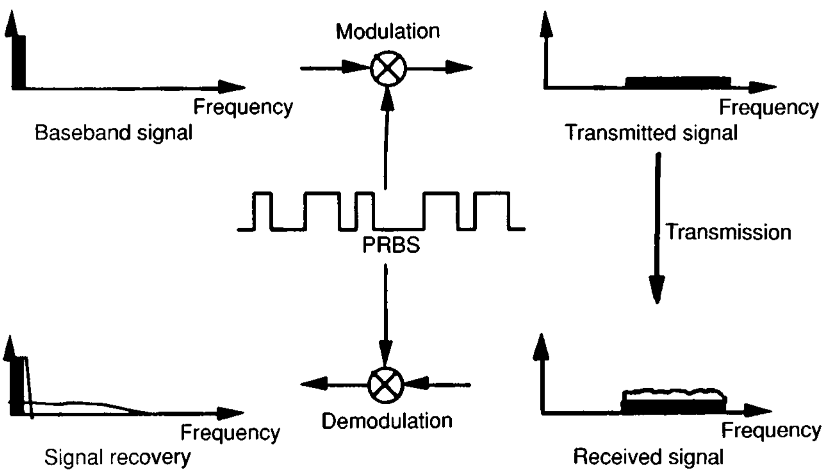


Fig. 1 CDMA Principles

in Figure 1. Received signals are then recovered by subtracting the corresponding PRBS codes. The bandwidth of the transmitted signal is "spread" by the higher bit-rate PRBS, which explains why this process is often referred to as Spread Spectrum Technology (SST). (0) This form of SST is known as "Direct Sequence". Another form of SST uses "Frequency Hopping", whereby a number of bits are transmitted at one frequency before hopping to a different pseudo-randomly selected frequency in the same band. Hopping typically takes place several times each second. The received signal is once again recovered through knowledge of the correct PRBS.

SST actually has its roots in military communications systems developed during World War II. Although complex, CDMA has a number of distinct advantages over other channel access schemes. Firstly, as individual channels are spread over the entire transmission spectrum, CDMA systems are virtually immune from narrowband interference. Secondly, it is very difficult to "decipher" a CDMA transmission without knowledge of its PRBS. Thirdly, very low power transmission is possible due to the robustness of the channel access scheme.

The primary channel access schemes are illustrated in Figure 2.

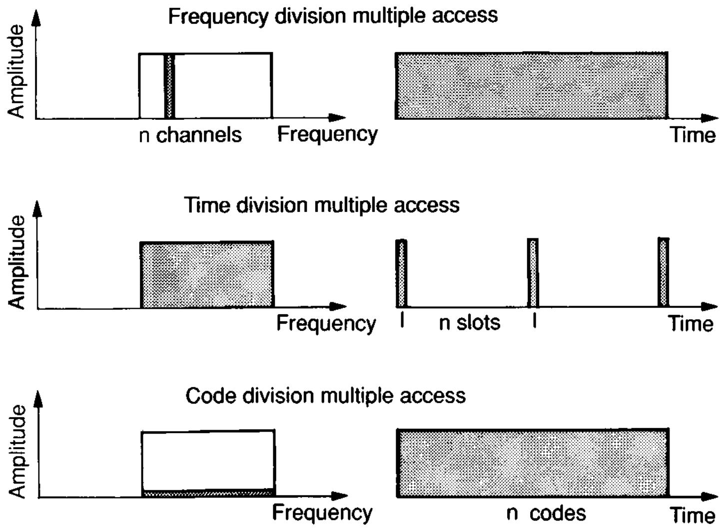


Fig. 2 Summary of FDMA, TDMA and CDMA

Careful choice of technology and channel access method will be required to support the target applications within the available bandwidth. It is important to understand how and why these choices are made and what tradeoffs result. An overview of the major Wireless Network systems and approaches is therefore presented below together with a comparison of their capabilities.

Wireless Networks may be divided into the following categories:
- LANs
- PBXs

– Public Telecommunications Services
– Mobile Data Networks

2.4 *Wireless LANs*

Wireless LANs are presently in their infancy but a number of products have been introduced since 1990, many based on SST radio and some using Infra-red. Wireless LANs are based on OSI principles and will eventually enjoy the benefits of OSI standardisation (see Section 5 below). Key issues to be resolved for Wireless LANs include the availability of adequate bandwidth (e.g. radio spectrum) and the provision of a high integrity data service (with a link bit-error-rate no worse than 1 bit in 10^9).

2.5 *SST Wireless LANs*

Shared sections of spectrum in the Industrial, Scientific and Medical (ISM) bands at 2.4 GHz and 5.8 GHz have been made available in a number of countries for low-power SST systems *within buildings*. Additional spectrum has been made available in the 900 MHz band within Canada and the USA. Complex CDMA schemes have so far been successfully implemented at data rates of up to 2 Mbit/s. The USA limit of 1 Watt peak transmission power provides an operating distance of up to 200 metres within a building; this is reduced to about 50 metres in Europe due to the recently imposed limit of 100 mW. It should be noted that the distance capability will be reduced when operating in the higher frequency bands due to increased propagation losses.

2.6 *High Performance Wireless LANs*

A new category of Wireless LANs is emerging for high speed operation – inspired, perhaps, by the fundamental limitations of CDMA. There are a number of initiatives associated with the use of 5 GHz, 17 & 18 GHz and 60 GHz frequency bands. The earliest initiative was from Motorola (1), who developed a 15 Mbit/s LAN system using 4-level FSK in the 18 GHz band. The Motorola system provides transparent support for 10 Mbit/s standard CSMA/CD but cannot be operated in the UK and a number of other countries due to the prior allocation of the 18 GHz band to fixed microwave links. The nearest available band within Europe is 17 GHz, and a request has been made to release some 200 MHz of spectrum for a high speed data service. The low transmission power and high frequency band used by the Motorola system will limit the distance capability to about 15 metres within a cluttered building environment. The next initiative, a project known as HIPERLAN (HIgh Performance European Radio LAN) comes from Europe and is targetting the 5 GHz band (more about this in Section 5 below). Looking ahead, Europe and Japan have identified the 60 GHz band as being appropriate for very high speed, short range LAN links.

2.7 Infra-red Wireless LANs

Infra-red light has been used to provide point-to-point links between buildings for some time. Highly focussed beams, spanning 100 metres or more provide a simple but effective form of communication. Point-to-multipoint schemes are made possible by either directed or diffused beams, as shown in Figure 3. Clearly, the diffused beam approach provides greater flexibility; however it also requires a significant level of optical power than a directed beam to cover the same distance.

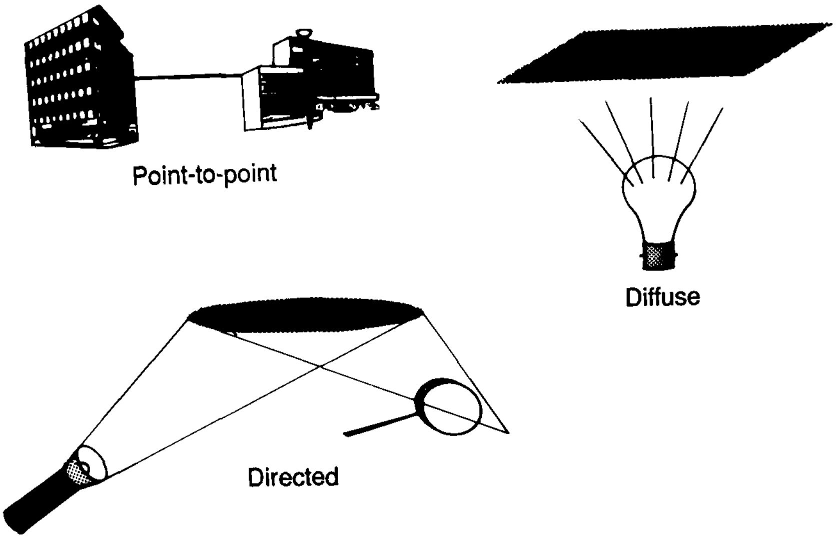


Fig. 3 Infrared modes of operation

Early development in Infra-red LANs date back to 1979, when IBM developed a prototype system at their Zurich Research Labs. The first InfraRed LAN operated at a bit rate of 125 kbit/s and used diffuse beams, taking advantage of the natural reflectivity of most ceiling materials, such as plaster and echo tiles. Distance and bit rate capabilities have been progressively improved over the years, but despite theoretical indications that 100 Mbit/s systems are possible for local user groups, no such capabilities have materialised. In fact, the majority of systems provide bit rates in the range 19.2 kbit/s to 1 Mbit/s. A detailed review of these developments is referenced (2).

A summary and comparison of the main Wireless LAN technologies is presented in Table 1.

Table 1 Comparison of Wireless LAN Technologies

	Radio LAN		InfraRed LAN
	Spread Spectrum	High-Performance	
Frequency Bands	900 MHz (USA), 2·4 GHz, 5·8 GHz	5·8 GHz, 17 GHz, 18 GHz (not UK), 60 GHz	—
Distance	50–200 m	15–50 m	10–25 m
Data Rate	2–10 Mbit/s	10–100 Mbit/s	1–20 Mbit/s
Security	Excellent	Moderate	Good
Integrity	Excellent	Good	Good
Cost	High & Falling	High	Moderate
Licensing	Unlicensed	TBC	Not Required

2.8 Cordless PBXs

Cordless (or Wireless) PBX attachment is emerging for application in the workplace, providing a range of up to 200 m (line of sight), reducing to around 50 m within a building. Unlike Wireless LANs, Cordless PBXs are voice-optimised with data added on at relatively low bit rates.

There is nothing unique about the PBX switch – "cordlessness" is simply added incrementally to service a limited number of lines to a local area or user group (via a local adaptor).

The main technologies for Cordless PBX are CT2 and DECT (Digital European Cordless Telecommunications).

CT2: CT2 is optimised to carry 32 kbit/s ADPCM speech over a full duplex digital channel in the 900 MHz band. A number of CT2-based products were announced in 1991, and some manufacturers have developed enhancements to provide extended services such as 2-way calling; these enhancements are known as CT2+ and CT3.

DECT: DECT is designed to carry a greater number of 32 kbit/s ADPCM speech channels within the TDMA framing shown in Figure 4. It is worth noting that the 420 bit DECT TDM frame includes an error checking field, which will facilitate the support of a high integrity data service. Depending upon the concatenation of time slots DECT will support a bit rate in the range 25.6 kbit/s to 153.6 kbit/s. The 1.8 GHz band has been allocated for DECT, and products are expected to emerge as early as 1993. Basic rate ISDN will be supported by DECT at some time in the future.

A summary of the main features of CT2 and DECT is provided in Table 2.

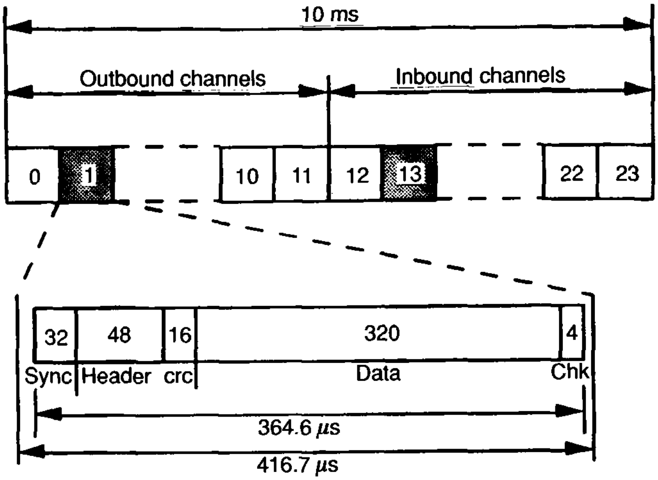


Fig. 4 DECT DMA Channel Access Scheme

Table 2 Comparison of Cordless PBX Technologies

	CT2	DECT
Frequency Band	900 MHz	1·8 GHz
Distance	50–200 m	50–200 m
Confidentiality	Yes	Yes
Data Support	Yes	Yes
Channels per Operator	40 max	120 max
Traffic Density	Low	10 × CT2
2-way Calling	with CT+/CT3	Yes
Initial Cost	£100 terminal	£400 terminal

2.9 Wireless Public Telecommunications Services

A number of Wireless technologies are expected to emerge during the 1990s. These public services are generally optimised for voice, although the support of data is gaining increased attention. Analogue cellular systems exist today in most developed countries, providing a mobile voice service (essentially vehicular) within major towns and cities, heavily-populated areas, and busy travel zones. Plans are now underway to replace these with digital cellular radio systems for both vehicular and pedestrian use. The key European digital technologies are Groupe Special Mobile (GSM), CT2 and Digital Cellular System DCS1800.

2.10 Analogue Cellular

Two systems are licensed to operate in the UK; TACS (Telecommunications Analogue Cellular System) and ETACS (Extended TACS). TACS and ETACS services are operated by Cellnet and Vodafone. Both UK services are overloaded, and the high level of congestion has driven down the quality of service to be barely adequate for voice use. This is understood to be representative of analogue cellular systems worldwide. It is possible to utilise the FDMA schemes for data by using FSK modems, operating typically at 2.4 kbit/s but possible at rates of up to 12 kbit/s through the use of more sophisticated devices; this will enable portable PCs and facsimile machines to operate remotely.

2.11 CT2

As previously stated, CT2 is designed to carry 32 kbit/s ADPCM speech over a full duplex digital channel. CT2 was actually developed as the basis for a low-cost personal (i.e. pocket phone) mobile voice service, planned for delivery in the UK via a number of consortia licensed to operate in the 900 MHz band under the "Telepoint" umbrella. CT2 supports a range of about 100 m from the nearest base station. CT2 technology does not facilitate 2-way calling, although this is possible by modification.

2.12 GSM

The majority of European PTTs are committed to GSM, which is planned to be the basis of a pan-European mobile service starting in 1993. GSM is oriented towards high-powered vehicular terminals and is forecast to exceed the use of existing European systems by 1995 (3). GSM is then anticipated to become a global service.

GSM will operate within the 900 MHZ band across Europe, thus facilitating "cross-border roaming". GSM employs a sophisticated and arguably expensive channel access scheme based on TDMA techniques; this has some similarity to DECT (see Figure 4). Speech is encoded at 13 kbit/s, with each speech channel processed within 20ms blocks (equivalent to 260 bits). Forward Error Correction (FEC) is also deployed within the GSM frames to recover from the effects of bursty, impulsive interference typical in a radio environment. The use of an FEC code make GSM a natural candidate for data and, indeed, a 9.6 kbit/s data service is planned as an integral service offering.

2.13 PCN/DCS1800

A pan-European Personal Communications Network (PCN) is planned for introduction in 1993. PCN will be based on DCS1800, which is a variant of GSM operating within the 1.8 GHz band and providing a range of between 500 m and 6 km (4). DCS1800 supports a bit rate of 9.6 kbit/s and a bit error rate of no worse than 1 bit in 100,000. One of the main advantages of DCS1800 is the increased traffic density arising from the use of the reduced base station service area. The likely success of PCN is the subject of much debate, with criticism of its high cost and complexity, and its direct competition with the planned Telepoint service.

A summary of the main features of Analogue Cellular, CT2, GSM and DCS1800 is provided in Table 3.

Table 3 Comparison of Wireless Public Telecoms Technologies

	Analogue Cellular	CT2	GSM	PCN/ DCS1800
Frequency Band	400 MHz 900 MHz	400 MHz 900 MHz	900 MHz	1·8 GHz
Distance	35 km	100 m	35 km	6 km
Confidentiality	No	Yes	Yes	Yes
Data Support	Limited	Yes	Yes	Yes
Channels per Operator	69	40	72*2	72*2
Traffic Density	Low 10% GSM	Good	Good	Excellent 3 × GSM
2-way Calling	with CT+/CT3	Yes	Yes	Yes
Cost	£20/month srvs+terml	£20/month srvs+terml	£90/month srvs+terml	£50/month srvs+terml

2.14 *Mobile Data Networks*

Data can be transmitted via Private Mobile Radio (PMR) networks using narrowband FM in the VHF bands. A number of public mobile data services have recently been launched to fulfil the increasing short-term requirement for portable data applications; these are limited however to national boundaries. Multi-national mobile data services are provided by satellite networks.

Private Mobile Radio Examples of PMR users in the UK are the major car rescue organisations, the AA and RAC. These users have optimised their national networks for data use by using tone modems in the analogue voice channels. This gives a relatively low data rate, but as data messages are generally more efficient than voice messages, the overall level of customer service is improved.

Public Mobile Data PMR systems require high initial investment due to the infrastructure needed to support a national network (base stations, trunks, etc). Users who are unable to justify the expenditure for PMR can instead use one of a number of emerging public mobile data services. The UK government recently licensed 3 companies to operate public mobile data services; RAM Mobile Data, Hutchison Mobile Data and Cognito. Additionally, a company called Paknet have UK licenses for connection to stationary and mobile terminals. All of these systems are packet-switched, cellular, wide-area networks, designed to cover the majority of the UK using some 200 base stations.

Satellite It is possible to operate a multi-national data service by using 1.2 kbit/s modems with the existing Immarsat A service (designed to provide a voice service at 1.7 GHz). Several satellite data services are planned.

Immarsat C is presently being implemented across Europe to provide a low-cost, compact data messaging service for sea and land mobile applications. Immarsat C operates on a packet-switched basis over the Immarsat satellite which is able to interface with a range of terrestrial messaging systems including X.25, voice band data and various EMAIL services. Looking to the future, a consortium led by Motorola is proposing to launch a network of 77 solar-powered satellites to provide a truly global voice and data service. This project, known as Iridium (which has an atomic number of 77), could become active as early as 1996.

A summary of mobile data systems is provided in Table 4.

Table 4 Comparison of Mobile Data Systems

	Data Over Cellular	PMR	Public Mobile Data	GSM	Satellite
Voice & Data	Yes	Yes	No	Yes	Yes
Data rate	2·4 kb/s	Low	6–10 kb/s	9·6 kb/s	0·3 kb/s
Coverage	National	National	National	European	Internatl
Terminal Cost	£1000	£1000	£500	?	£2000

3 Frequency Management

As radio is the most appropriate medium for all wide-area and the majority of local-area Wireless Networks (as discussed in Section 2), it is absolutely essential that radio spectrum is managed effectively. This is accomplished on a worldwide basis by the various regulatory bodies, not however without difficulty, and not with great speed. The key issues and considerations associated with radio frequency management are summarised below, together with a synopsis of the regulatory framework that currently exists.

3.1 Choices and Considerations

3.1.1 Operating Range The operating range of a radio network will largely be determined by transmission power, the propagation characteristics relating to the frequency band used, and the operating environment. The transmission power appropriate for point-to-point terrestrial links and portable hand-held terminals, for example, will differ significantly. Portable hand-held terminals will have low transmission power due to battery conservation and health & safety issues (see below), whereas a point-to-point terrestrial link could be operated safely and easily at very high power levels. As mentioned in Section 2, radio propagation will change with frequency; with the higher frequency bands less able to penetrate dense structures such as buildings. Frequencies above 2 GHz are generally recognised as being more useful for line-of-sight applications and may have difficulties when used for large zone local networks. Cluttered operating environments such as city centres and within-building will reduce the effective range of a microwave radio network.

3.1.2 Health & Safety At frequencies above 30 MHz there is a particular concern relating to the heating of body tissue by radio frequency energy (5). The human organs identified as being most prone to rf heating effects are the eye and brain. Exposure time comes into the equation due to the thermal mass of the human body and organs. Strictly imposed limits are therefore issued by the regulatory bodies in order to protect users; these are especially applicable to hand-held mobile telephones which can couple a high proportion of their transmitted energy into the eyes and brain due to the close proximity of the antenna. Some systems have a built-in timer which automatically reduces the transmission power after several minutes – CT2 is an example of such a system.

3.1.3 Bandwidth As radio spectrum is a precious, limited resource its allocation to the various operators and user communities is, without doubt, the most fundamental consideration. Allocation is managed by the regulatory bodies (see below). There are a number of factors to be considered in matching traffic requirements to radio spectrum; the key points being:

a) The business case to support frequency allocation for a particular service, or set of services (6). This is often the first point of "negotiation" in the bid for spectrum, and often data will compete with speech applications for the same spectrum!
b) The total traffic bandwidth required, and whether this requires the radio spectrum to be contiguous or allows it to be fragmented.
c) The efficiency of use arising from the modulation technique and channel access scheme employed (see Section 2).
d) The re-use factor associated with cellular systems, i.e. the smaller the cell, the greater the re-usability of radio spectrum used.

3.1.4 Licensed vs Unlicensed Operation Public and private wireless network system operators find a number of benefits from running a licensed service; this will give them dedicated bands, with the protection of interference from other users. Wireless LAN users, on the other hand, will generally feel that the administration associated with licensing is a burden they do not wish to carry. It is therefore generally felt that "unlicensed" operation is more appropriate for this application. A number of unlicensed frequency bands have already been made available by the regulatory bodies – notably the ISM bands; however the bandwidth available will not support the more demanding LAN applications, and the prospect of "overlapping" Wireless LANs raises a number of concerns. Some solutions for the management of shared spectrum are considered in Section 4 below.

3.2 Regulatory Framework

The globe is divided into 3 distinct regions for the purpose of radio planning. These zones are defined by the ITU (International Telecommunications

Union) which organises the World Administrative Radio Conference (WARC) as broadly:

Zone 1: Europe, Scandinavia, Russia and Africa;
Zone 2: North and South America;
Zone 3: Australasia and Far East.

Europe is beginning to see the allocation of pan-European spectrum for some applications, but a large number of allocations are unlikely to become common between countries. It is fortunate that a number of Wireless Telecommunications technologies have become the subject of intense debate throughout Europe and that common standards and frequency allocations are being agreed.

The radio regulatory authority within each individual country is responsible for the allocation of spectrum within that country. European harmonisation is provided within CEPT (Conference of European Postal & Telecommunications administrations), which is understandably influenced by the CEC. However, ETSI (European Telecommunications Standards Institute) is the

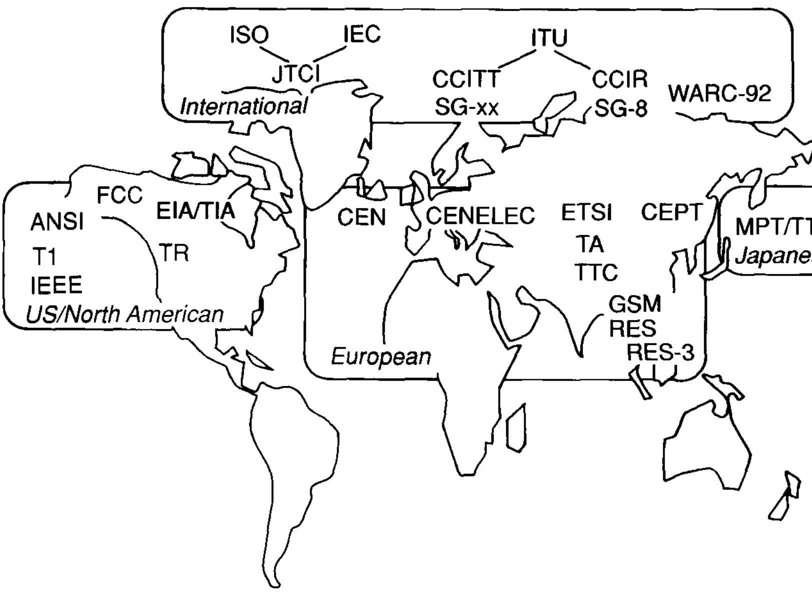


Fig. 5 Wireless Standards and Regulatory Bodies

only body which can define and issue pan-European Standards. Figure 5 presents the major radio standards and regulatory bodies worldwide, and a detailed review is referenced (7).

3.2.1 Licensing Regimes A number of license types exist depending on ownership and applications. This will provide control over the major sectors, and will assist in spectrum assignment. The main types of license and applications are as follows:-

Operator's License – required for fixed and mobile public telecommunications services, traditionally supporting speech services but increasingly data (e.g. GSM). This license would provide legal protection from interference by other users.

Site License – applicable to radio communications equipment used within a single site. This type of license would prevent the use of approved apparatus off-site and is therefore thought not to be appropriate for some types of application (e.g. ad hoc LAN). This license would also provide protection from interference by other users.

Owner's License – to cover special requirements such as private point-to-point links, etc. Once again, this license would provide protection from interference by other users.

License Exempt – to cover operation in unlicensed bands (e.g. ISM bands). This option requires that interference *to* licensed users shall be kept to an acceptable level but provides no protection from interference *by* other users. Product Type Approval is required and within-building restrictions usually apply. License Exempt operation is currently favoured for Wireless LANs.

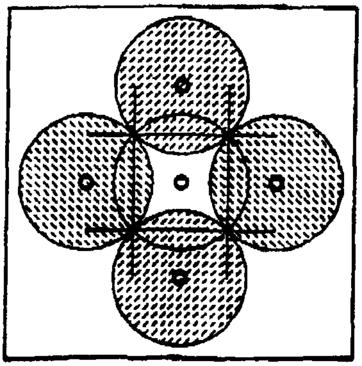
Class License – this applies to products (Type Approval), and is performed by, or on behalf of, the national regulatory body to ensure compliance in areas such as frequency stability, power levels, interference, etc.

4 Architecture

There are many complex and interrelated factors associated with Wireless Network design. An attempt is made below to summarise the key architectural considerations, although a number of these items are clearly related to Frequency Management.

4.1 Cell Size

Factors affecting the operating area covered by a transceiver are listed above in Section 3. If it is possible for all participants to operate within a single cell (i.e. they remain within a common operating range), then the only additional consideration will be how to share the available bandwidth among the users. Whilst it may be very tempting to design such a system, the single cell (or monozone) approach has a number of limitations. These limitations are overcome by splitting the total service area required into multiple cells, hence the term – cellular. If we consider the effects of reducing the cell size as shown in Figure 6, the gains in relative data rate, transmission power and total bandwidth capacity are presented in Table 5 (8). Figure 6 utilises



Detail of circular coverage overlaps from four adjacent coverages using a range sufficient to reach a diagonal corner

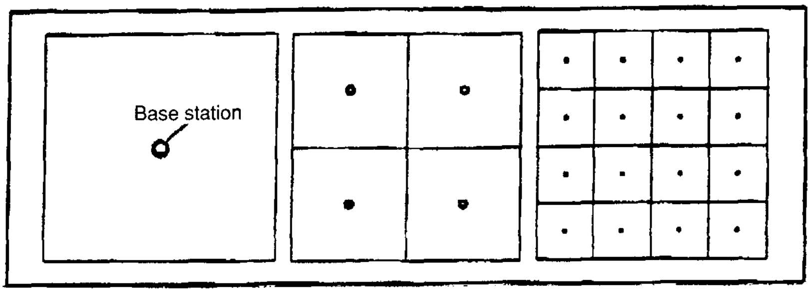


Fig. 6 Complete layouts for the Access-points with Reach Ratios 1:0.5:0.25

base stations which have overlapping coverage, as shown. It can be seen that the use of micro-cellular systems provides a number of benefits:-

Spectrum Re-use Radio spectrum may be re-used when interference becomes acceptably low – this is judged to be possible at approximately 3X the operating range of a base station (i.e. every 4th contiguous cell) (8). Spectrum re-use is a significant incentive to deploy cellular systems.

Bandwidth Density Fewer users share the total available bandwidth within a micro-cell, hence the bandwidth density will be higher. This could be an important consideration if the user density and application bandwidths are high (as they will be with a LAN).

Power Consumption Transmission power will reduce significantly with cell size. This will contribute greatly to the complexity, size and cost of devices, and will be key in setting the operating times of battery-powered, portable units.

Table 5 Scaling factors for time dispersion limited wireless access-point grid

Grid Pitch DIM "D"	Range ×D of 1	Area Per Access-PT	Relative Data Rate	REL Power Required	Capacity Per MHZ
1	0.7	1	1.0	0 dB	1.0
2	1.4	4	0.5	+6 dB	0.125
4	2.8	16	0.25	+12 dB	0.0156

4.2 Topology

Most Wireless Network configurations will be hierarchical, based on trunked cellular structures. In the wide-area this will comprise an array of base stations which are interconnected via the public telecommunications trunk network. In the local-area it will often be the interconnection of a number of base stations via a private building or campus backbone. It is worth noting that the majority of trunk and backbone systems will be cabled media rather than wireless, with wireless providing the final connection from a predominantly cabled infrastructure (9).

Some Wireless LAN applications may require a peer-to-peer configuration, an example of this being "ad hoc networking". In practice, a group of portable PCs could be networked for meetings, conferences, field activity, etc. This is a real multi-vendor environment which does not favour the use of base stations.

4.3 Protocol Stacks

4.3.1 Wireless Networks are generally seen as part of a larger, hybrid infrastructure which is supported by a high-speed trunk system (as discussed above). It is therefore generally felt that the Wireless Network should be structured according to the 7-layer OSI model, architecturally limited to Layers 1 and 2. This is generally the case (10), however there are some notable additions:-

4.3.2 Integrity The quality of service provided by a radio link is judged to be adequate for speech application but is not high enough to support many data services, particularly datagrams. Bit-error-rate requirements for 8802-LANs (Token Ring, Ethernet) are no worse than 1 bit in 10^9 at the transmission level, whereas the typical bit-error-rate observed within a radio channel is 1 bit in 10^3. This shortfall may be resolved in 2 ways; by using an FEC (Forward Error Correction) function at Layer 1, or by Transport Layer acknowledgement, or both.

4.3.3 Confidentiality Radio is a natural broadcast medium and therefore confidentiality will be an important requirement. The emerging Wireless Telecommunications services (GSM, PCN) have taken this into account, but Wireless LANs have yet to resolve this issue. Wireless LANs could use spread spectrum techniques (refer to Section 2), however other levels of protection may also be required. Levels of confidentiality being considered for use with Wireless LANs include Layer 1 encryption and/or the IEEE 802.10 Standard for Interoperable LAN Security (SILS) (11).

4.4 Management

There will be a number of management features unique to Wireless Networks; one will relate to the operation within cellular structures, another to the use of shared radio spectrum.

4.4.1 Roaming This term applies to mobile terminals, and their ability to secure continuous service when traversing multiple cells. This is prerequisite for all mobile services and is made possible by system management which will monitor the location and status of each participant and thus provide 2-way calling and "seamless handover" between cells. The requirement for roaming is less obvious for Wireless LANs but is being considered by the standards committees.

4.4.2 Coexistence The management of spectrum allocated for any particular service will, in practice, necessitate dynamic bandwidth assignment to participants. There is nothing unusual about this, however additional consideration is required when multiple user communities attempt to access the same spectrum (e.g. ISM bands). The latter is a particular problem with Wireless LANs. This problem may be resolved by the use of an appropriate "etiquette" which establishes the rules as part of any initialisation procedure and continuously monitors the status of the network. Such an "etiquette" is under development within the Wireless LAN standards committees, and is being viewed as a distributed management entity.

5 Standardisation

There are several candidate standards and technologies vying to champion the cordless PBX. CT1 is a 1980s analogue technology and is not really appropriate for cordless PBX applications, other than perhaps single exchange line systems. CT2 was developed for domestic application but has since been adopted by ETSI for Telepoint services, and many PBX manufacturers are now adopting this technology for their cordless PBX products. CT2 is digital and is best suited to low-density attachment. It does not allow hand off between base stations, although Northern Telecom has developed an enhancement that does (CT2-plus). CT2 cordless systems do not possess an inherent data capability. CT3 is a proprietary pre-DECT technology developed by Ericsson, however ETSI has recently rejected this, and no other manufacturer is looking at it. The most futuristic and capable standard which is presently being agreed within ETSI is DECT (Digital European Cordless Telecommunications). DECT is optimised for cordless PBX application and is able to serve a greater density of users than CT2. DECT will eventually support Basic Rate ISDN, which has obvious implications as a future office technology.

European standards for PCN/GSM have been agreed by ETSI, and products are planned for release in 1993.

Development of Wireless LAN standards commenced within IEEE 802.11 in 1990 and this work is well supported. This group will generate a standard Medium Access Control (MAC) protocol optimised for cellular radio propagation plus multiple transmission (Physical Layer) standards operating at bit rates in the range 1–20 Mbit/s, although 50 Mbit/s is understood to be possible. Due to the poor integrity of the physical radio channels a Forward Error Correction protocol will be required – this will necessitate the use of small data packets. Initial standards are expected to become technically stable circa 1994, possibly with a high bit-rate version following in 1996 (12).

IEEE 802.11 is giving very little emphasis to Infrared.

ETSI Committee RES 10 is developing a high-speed radio LAN known as HIPERLAN (HIgh Performance European Radio LAN). This project commenced in 1992 and is well supported by industry, however an approved standard is not expected to be available before 1995. ETSI has identified the task of spectrum allocation as being fundamental, and is working with CEPT to secure a sizeable slice of the 5 GHz band (100–150 MHZ) before it commits resources to actual standards development. ETSI is also discussing the availability of an appropriate licensing regime with the European regulatory bodies. These manoeuvres should add to the likely success of this European project.

6 Applications and Business Benefits

6.1 Who Need Wires?

LANs have become the de-facto standard for desktop data connectivity. By the mid-1990s it is expected that the majority of desktop data equipment will be attached to a LAN. Structured cabling has evolved to satisfy the requirements for flexibility, expandability, cheapness and performance. While well-suited to modern office environments, it has its drawbacks.

Cables and wires are dangerous and unsightly in working environments. People trip over them, they are pulled and snagged, they are invariably the wrong length and they continually get entangled with other cables and tie themselves in knots. There are never enough sockets and they are invariably of the wrong type. Cables fill ducts and voids and once in, they stay in, because, as they aren't labelled and everyone who knew what they were left years ago, no one dares remove them.

Structured cabling has come a long way towards alleviating this rather nightmarish scenario, but there is no doubt that were wireless technology as efficient and cost-effective as cable at carrying large quantities of information at high speeds, IT cabling would have been dead and buried long ago.

But it isn't, and for the majority of business applications structured cabling is, and will for the foreseeable future continue to be, the most cost-effective means of providing ubiquitous networking. With a plethora of connections supported – practically all proprietary protocol and de-facto standards, as well as international standard LANs – structured cabling is in a benevolent cycle that shows no signs of abating. It would take an extremely courageous Facilities Manager to embark today on a networking strategy that totally excluded cable.

Nevertheless, the wireless scenario is an extremely compelling one, and its attractions will inexorably drive the technology, performance, standards, bandwidth availability and international harmonisation into serious competition with cabling in an ever-increasing range of circumstances. Wireless LAN applications are expected to be niche until standards emerge around 1994–5, after which time WLANs will enable new applications and also begin slowly to erode the cable market. Early applications will tend to support mobile staff, "problem areas", and rapidly reconfigurable working positions.

The most likely short/medium scenario foresees:

- pockets of wireless LAN overlaying conventional cabling to provide instantaneous LAN connection for laptop PCs and mobile personnel
- wireless increasingly used in areas difficult or expensive to cable or with short life expectancy
- increasingly widespread use in niches such as supermarkets, factories, hospitals, university campuses where specific benefits are worth the extra outlay
- wireless for standby in case of cable plant failure

6.2 Cable versus Wireless

6.2.1 Cost At present, cost is only one of a number of issues affecting the choice of cable versus wireless. But as wireless standards, wavebands and wireless LAN technologies become established in the latter half of the decade, cost, related to bandwidth, will increasingly become the determining factor. Then the higher initial costs of a wireless solution will need to be offset against the costs, timescales, environmental disadvantages and general disruption of cable installation and cable-related moves and changes. Today, structured cabling typically costs around from £50 per installed point compared with upwards of £400 per wireless LAN node.

The balance swings in favour of wireless:

- as the cost of installing cable increases (cost can run into thousands of pounds per cabled point in listed buildings compared to around £70 per point for a normal building)
- as the cost and/or duration of moves and changes increases (estimated at £10–50 per move with structured cabling, but rising to hundreds of pounds and requiring careful planning if comprehensive flood-wiring is not installed)
- as the frequency of moves and changes increases (generally between once every 6 months to 5 years depending on industry and environment)
- if data equipment is not uniformly and frequently distributed throughout the space

6.2.2 Flexibility The overriding characteristic of wireless is its profound flexibility – equipment can be sited and moved anywhere without reference to cabling outlets. The greater the value of this flexibility to an organisation, the sooner is a wireless solution likely to be judged cost-effective.

6.2.3 Reliability Reliability approaching 100% is paramount for information networks today for the effective running of modern business. Wireless technologies can now meet the exacting standards demanded by modern organisations reliant on Information Technology and networking, leaving little to choose between wireless and cabled networks as far as reliability is concerned. Indeed, being less prone to cable snagging and accidental damage, wireless networking may well have the edge.

6.2.4 Security As with reliability, security is another area in which the general perception has been that wireless technology fails to meet the exacting standards demanded by modern organisations, but again the reality is that wireless networks are quite as secure as cabled networks – using specialised techniques such as frequency-hopping in addition to the common encryption and other forms of protection.

6.2.5 Standards It is in the area of international standards that wireless begins to fall down. Whereas there are well-accepted international and industry standards for both structured cabling infrastructures and the LANs running over them, international standards for wireless technology are not expected to be in place before 1995 at the earliest – therefore early implementors will find themselves tied to particular products and suppliers with little interworking capability. Users like freedom of choice and a feeling that their investment is protected, and don't like multiple standards – witness the failure of the original CT2 venture (Telepoint, Phonepoint, Zonephone).

Additionally there is little international agreement in the allocation for general-purpose use of the radio spectrum, most of which is allocated to military use. This again means little international harmonisation or product interoperability.

However, standards bodies and suppliers are well aware of the need to move quickly, and the European Telecommunications Standards Institute (ETSI) has announced its intention to produce a draft "HIPERLAN" wireless LAN standard that can effectively deliver 10–20 Mbits/sec to the desk by end-1994, with full approval by end-1995.

6.2.6 Bandwidth Bandwidth is the other principal area of shortfall of wireless compared to cabled networks. Great strides have taken place recently in the extraction of ever-increasing performance out of humble twisted-pair wiring; high-grade UTP cable used in structured cabling systems now supports 20 Mbits/sec, with 100 Mbits/sec and more anticipated for the near future, whereas wireless is struggling to exceed 2 Mbits/sec. Consequently the cost of achieving a given throughput level is considerably less for cabled networks than for wireless, and this differential is likely to continue as the performance of both increases.

Nevertheless, a strategy including wireless "nodes" at strategic positions to provide local distribution, wholly or partly replacing floor-cabling, in a network backbone-cabled with copper or fibre, is very attractive, provided that the higher-level interfaces (Ethernet, Token Ring etc.) are maintained, and that bandwidth, at least in the short term, is not a major issue. The viability of this approach is determined by the cost-benefit of the additional flexibility, and the costs associated with cabling the environment, as outlined above.

6.3 General Wireless LAN Scenarios

6.3.1 Office Environments For ordinary office environments, a cabling solution will be the most cost-effective for general use, offering the best cost-performance at any given point in time. For new or refurbished buildings for normal office use, structured cabling is the recommended approach for the foreseeable future.

However, there are several circumstances in which wireless LANs will rapidly gain credibility.

6.3.2 The Laptop Explosion By 1995, half of all PCs sold will be notebooks or laptops. Their proliferation, coupled with widespread Ethernet and Token Ring networking, will combine to create an expectation of LAN connectability from wherever one happens to be working – certainly in any part of a building or campus; also in the car, in the train or working at home (13). Data messaging will be required between laptop and hand held PC devices, and there will be a need to interconnect portable PC devices to the corporate network when used within a general office environment.

Wireless overlays are the only effective means of providing such global ad-hoc capability, but attachment costs will need to be low for mass adoption.

However, reduced performance compared with cabled LANs will probably be acceptable at least in the short term.

6.3.3 Multi-Purpose Areas Areas not originally intended for use as office environments will probably not have been fitted with structured cabling. Examples are:

- Restaurant and refreshment areas
- Lobbies, foyers, halls, entrances, gangways and general circulation areas
- Exhibition areas
- Storage areas
- Recreation areas
- Training rooms
- Boardrooms
- Conference rooms
- Computer rooms
- Individual offices

Networking information equipment with cabling in such areas can often be problematic – for example tills and terminals in restaurants and lobbies, computerised equipment in storage areas, equipment used for demonstration, training or promotional events. Change of use for temporary or permanent office accommodation may also be better networked by wireless LAN rather than cabling.

6.3.4 Mobile Personnel and Portable PCs People whose jobs cause them to be regularly on the move are making increasing use of portable terminals – examples include salespeople, site or facilities management personnel, service engineers, security operatives. Their needs for network information access or data entry will be most effectively satisfied by wireless (14).

6.3.5 Mobile Units Wireless technology is undoubtedly the best solution where network access is required from mobile offices or other vehicles – for example positioned near buildings for security purposes or exhibitions.

6.3.6 Problem Environments – Areas Difficult or Expensive to Cable In general, older buildings are likely to contain areas where comprehensive cabling is likely to be prohibitively expensive or disruptive to cable, and which are therefore ideal candidates for wireless LANs. False floors may be difficult or impossible to provide, ducts and voids may be small or non-existent, and internal walls solid and thick. The cost of providing flood-wiring could well be prohibitive, making the subsequent costs and logistics of moves and changes expensive. Wireless networking, perhaps coupled with a cabled backbone network, will often be the most cost-effective solution.

6.3.7 Listed Buildings Listed buildings are an extreme example of Problem Environments – not only may cabling be difficult, disruptive, time-consuming and expensive, it may not even be permitted! Alternatively, compliance may be very costly. Wireless is the obvious solution.

6.3.8 Outside Areas Wireless LANs can provide cover in areas where cabling would probably not even be considered on account of expense or installation problems:

- Car parks – control points and booths
- Security posts
- Garages
- Filling station booths
- Building or campus peripheries
- Sports areas
- Outdoor or under-canvas exhibitions

6.3.9 Outbuildings Buildings remote from the main site can often be linked to the main network more efficiently by wireless technology than by cable, probably using direct microwave or infra-red links rather than a wireless LAN per se. Examples include temporary buildings, mobile units and security posts.

6.3.10 Temporary Buildings An extreme example of an outbuilding is a temporary building – Portakabin or similar. Cabling may be difficult because of the design of the building, or too short-term to be cost-effective. Wireless LANs are the solution.

6.3.11 Temporary Networks Wireless LANs can provide instant facilities for internal or external meetings, conferences, showrooms or exhibitions. External meetings and conferences represent a genuine multi-vendor environment, and clearly will not be possible until standards are mature.

6.3.12 Short-Term Provision Wireless LANs will be invaluable to provide networking facilities for short-term building occupancy, to protect investment in LAN hardware, e.g. to give temporary cover in parts of buildings scheduled for refurbishment in the near future, where network connection is required but cabling would clearly not be cost-effective.

6.3.13 Emergency cover Wireless LANs are likely to play an increasing part in the provision of instant resilient back-up systems to provide at least partial emergency network connection in the event of failure in the cabling system – to provide an instant network for example where a cabling node or trunk has been damaged by fire, accident or wilful damage.

6.4 Industry-Related Application Areas

6.4.1 Retail Supermarkets and department stores are inherently unsuited to cabling as their terminals are generally clustered at checkout points rather than evenly spread around the floor area. Checkout positions, particularly in multi-purpose department stores, need to be relocated or altered in layout or size to respond to seasonal fluctuations or changing fashions, and speed of response to changes in demand are crucially important to the slim profit margins in the industry.

Marks and Spencer pioneered the use of wireless LANs in this field, with an ICL solution using Radiolink, a product based on spread-spectrum technology.

Specific applications include:

- Rapid reconfiguration of POS terminals to enable unplanned changes in stores which are used 6–7 days a week, and where manual labour costs are high
- Ideal network medium for shelf-edge price displays
- Shopping trolley "route" monitor to determine optimum goods presentation
- Stock/price enquiry terminals (hand held)
- Stock control; for goods-in/goods-out/stock-taking (hand held)
- Remote diagnostics for dispensing machines for food, drink & other goods (e.g. needs filling/emptying)

6.4.2 Finance Wireless LANs are likely to be of particular relevance to banks and dealer rooms. Many banks are sited in old buildings difficult to cable effectively; attended or customer-operated terminals (ATMs) are often clustered at service points. However, in favour of fixed cabling, layouts tend not to change very often.

Dealer rooms on the other hand are crammed with as much terminal equipment as can possibly be squeezed in, changing continuously. Designing cabling for such environments is nightmarish – how much, what type, where; clearly an excellent opportunity for wireless, given sufficient bandwidth.

Other applications include:

- Networked, hand held PCs for stock trading
- "Instant" LANs and short-term networking facilities for auditors and ad hoc workgroup activity

6.4.3 Transport Potential applications for wireless LANs exist in stations, airports, customs terminals as well as on ships. All have large multi-purpose retail, commercial and business areas with mobile personnel and a need for flexibility in response to seasonal variations, changes in security requirements and change of use.

6.4.4 Manufacturing and Commercial Opportunities for wireless LANs abound – on factory floors where process control equipment is increasingly networked and cabling is inefficient and hazardous, in automated warehouses and storage areas, plant and equipment areas.

Specific applications:

- Networking industrial automation mobiles (robots) and process-control equipment
- Tracking of goods in progress (using hand held data entry terminals)
- Automatic tracking of transport vehicles (using microcells)
- Temporary network facilities for new building sites, including voice, text and image services

6.4.5 Security Police, security and defence organisations have always been foremost in the demand for wireless networks to support mobile and field personnel requiring ever more comprehensive access to information bases. The recent upsurge in demand for security monitoring and control in public places will extend the applicability of wireless technology both for networking fixed surveillance and monitoring equipment, and for supporting mobile personnel.

Physical location of portable PC devices anywhere within the granularity of the operating zone (which could be as little as 10 m within a building or as great as 25 km outside) would enable improved security cover for both people and equipment.

6.4.6 Utilities Specific applications include:

- Data entry and enquiry facilities for the many meter readers and field support personnel associated with the utilities
- Inventory management for the high level of stock held and transacted by many of the utility companies (via hand held terminals)
- Field access to centrally-stored configuration plans, with the possibility of co-ordinated access to overlaid utilities (e.g. underground gas, electricity, water, sewage & telecoms systems)
- Telemetry for remote parts of the utility infrastructures, including meter reading (and therefore reducing the human element, and its cost, significantly)
- Security monitoring and surveillance inside and outside buildings
- Showroom and shop opportunities as for Retail

6.4.7 Public Service Industries

- Remote salesforce support, including mobile access to product databases, inventory checking, order entry and financial transactions. Smart

laptop PCs running "multimedia" applications as a sales aid (linked to central database).
- Remote, specialist salesforce support, with mobile access to complex modelling tools to support financial investment, insurance, or other purchases which have many variables.
- Data entry (via a compact hand held terminal) to streamline ordering in restaurants, check-in process for hire car returns, etc, including mobile credit card authentication and charging

6.4.8 *Commercial Sales and Customer Services*

- Remote/mobile access to expert diagnostic system and personnel, extending the capability of the person in the field
- 2-way messaging to facilitate data entry (e.g. fault fix, spares, time charging) and assist in dynamic scheduling of field personnel
- Tracking of field personnel and vehicles and providing routing directions

6.4.9 Local and Central Government Applications include:

- Tax, rates, unemployment etc offices with clustered and mobile terminals
- Field access to utility maps and plans
- Field access to demographic, housing and personal information
- Security and surveillance
- Plant and equipment rooms
- Education and leisure centres
- Vehicle and field personnel tracking

6.4.10 Health Cabled networks in hospitals may not only be inefficient in wards and other areas where hospital staff require access to patient information from anywhere within the building, but the presence of cables in such areas may actually pose serious danger to people and equipment. Mobile equipment will increasingly require network connection – cost-effective wireless LANs are the solution.

Applications include:

- Data entry facility for Social Services personnel, GPs and other healthcare professionals who are required to be highly mobile, with tight schedules, and often little time for data entry (e.g. medical records)
- Improved access to medical records (case histories, X-ray images, etc) for healthcare professionals, who are required to be mobile for a high proportion of their time (both locally and remotely)
- Operational support for ambulance crews, etc.

6.4.11 Education Universities, polytechnics and schools all need to be able to network equipment from anywhere within their buildings or campuses.

Applications include:

- File/database/library access, within the classroom and campus, and also on a wide-area basis (e.g. Open University). Largely based on laptop PCs gaining access to text and image services.
- Interactive learning, largely based on laptop PCs running "multimedia" applications (i.e. text, image, voice and video). Geographical scope ultimately as above.
- Field study support, with remote database access and local networking facilities

6.4.12 Defence All services require extensive and comprehensive field networks for mobile equipment and personnel – an obvious application of wireless LANs able to meet the security criteria.

6.5 Cordless PBXs

The main take-up of cordless PBXs will be within the service industries (greatest need for mobility/rapid response), and also on larger multi-building sites. The major application of cordless PBXs will be voice – data will be the poor relation, and will begin to gain support beyond 1995. No territorial variations are known.

One of the main drivers for cordless PBXs will be the need to improve the efficiency and effectiveness of *locally* mobile personnel. Costs will determine the rate of take-off, and synergy between different types of service will also help; an example of such synergy is the ability of CT2 terminals to operate across Telepoint, domestic and cordless PBX domains. The latter is expected to reinforce the success of CT2 as the initial standard, and a number of major suppliers plan to launch CT2-based products in 1992. 1992 Costs of CT2 products are in the region of £200 for a handset and a further £400 for a share of the base station (15), reducing to about £200 total by 1995. Initial DECT-based products are expected to emerge in the 93/94 timeframe. Forecasts for initial costs of DECT attachment point to it being roughly twice the cost of CT2. It is presently difficult to assess how soon these costs will fall. CT1, CT3 and possibly CT2-plus are expected to have a limited future.

Competitive costs associated with structured cabling will hold off any *general* take up of cordless PBXs (note that a wired handset and PBX cabling currently costs circa £100).

Market forecasts indicate that GSM is expected to compete effectively with existing analogue cellular systems, matching its use by 1995 and reaching a penetration of 60 subscribers per 1,000 (3) shown to be grossly optimistic to date, with some UK consortia recently claiming to have financial difficulties.

Market forecasts indicate that the penetration of PCN will match Telepoint/CT2 in 1997/98, with a dramatic take-off soon afterwards.

The implementation of GSM has the commitment of PTTs in the EC, EFTA and several other countries. Its rate of take-up will be influenced by the need for high-speed/large area cross-border roaming with high quality of service (including seamless handover between cells). GSM will no doubt be competitively priced against existing services, however analogue cellular systems may continue for much of the 1990s within less populated areas. PCN and Telepoint will compete for the same market, however PCN possesses a number of technical advantages over Telepoint regarding the support of data (integrity and throughput). Unfortunately, PCN is expected to be significantly more costly to establish, with estimates of £1–2Bn to set up a PCN system within the UK (compared with £70M to set up a network of 25,000 base stations). The latter supports the expected late take-off of PCN and suggests that Telepoint may experience a useful window of opportunity.

References

0 SCHILLING, D.L. et al.: Spread Spectrum for Commercial Communication. *IEE Communication Magazine*, April 1991, pp. 66–79.
1 BUCHHOLZ, D et al.: Wireless In-Building Network Architecture and Protocols. *IEEE Network Magazine*, November 1991, pp. 31–38.
2 BARRY, J.R. et al.: High-Speed Nondirective Optical Communication for Wireless Networks. *IEEE Network Magazine*, November 1991, pp. 44–54.
3 DETTMER, R.: GSM – European Cellular Goes Digital. *IEE Review*, July/August 1991, pp. 253–7.
4 POTTER, R.: Personal Communications for the Mass Market. *Telecommunications*, September 1990, pp. 79–80.
5 GRANT, E.H.: Biological Effects and Health Hazards of Exposure to Radiowaves and Microwaves. *The Radio Spectrum Management Conference, CommEd*, March 24, 1992, London.
6 COLLINS, H.: The Economics of Spectrum Allocation. *The Radio Spectrum Management Conference, CommEd*, March 24, 1992, London.
7 WILKUS, S.: Standards and Regulatory Aspects of Wireless Local Communications. *IEEE 802.11 Standards Committee Working Paper* 91–63, May 1991.
8 RIPINSKI, C.A.: Wireless System Architecture – Major Choices & Considerations. *IEEE 802.11 Standards Committee Working Paper 90–18*, November 1990.
9 FLATMAN, A.V.: Universal Communications Cabling – A Building Utility. *ICL Technical Journal*, Vol. 6, no. 1, pp. 117–136, 1988.
10 RIPINSKI, C.A.: Wireless LAN Architecture – Topology & Protocol Stacks. *IEEE 802.11 Standards Committee Working Paper* 91–21, February 1991.
11 IEEE 802.10 Standard for Interoperable LAN Security (SILS) – draft.
12 HAYES, V.: Standardisation Efforts for Wireless LANs. *IEEE Network Magazine*, November 1991, pp. 19–20.
13 BIBA, K.: Wireless LAN Market Analysis and Forecast, *IEEE 802.11 Standards Committee Working Paper* 92–20, Jan '92.
14 FARWELL, C.: Business Benefits of the Cordless Office. *The Cordless Office Conference, CommEd*, May 24, 1991, London.
15 ZIZIROS, M.: Overview of the Cordless Office Market. *The Cordless Office Conference, CommEd*, May 24, 1991, London.

Biographies

Alan Flatman

Alan Flatman received a first-class honours degree in Electronics Engineering at the North Staffordshire Polytechnic in 1971, and later gained a Ph.D in Electronics and Computing in conjunction with CNAA and ICL. He joined ICL formally in 1975 and has performed a number of roles associated with communications products and technologies. Since 1981 he has been responsible for LAN technology, standards and technical strategy, and has been an active participant in UK, US, European and International standards committees. He presently works within ICL's OPEN-*framework* Division where he is responsible for LAN Architecture and Conformance. He has over 20 publications and is a Fellow of the Institution of Electrical Engineers.

Paul Weiser

Paul Weiser is ICL's LAN Marketing Manager in Network Products Unit, Logistics Division, with responsibility for strategy, selection, marketing and exploitation of ICL's LAN product portfolio.

With 25 years in the computing industry, Paul Weiser has extensive experience of all its aspects. He has held a variety of Marketing, Consultancy and Project Management positions covering the whole breadth of computing from mainframes to office systems and networks.

A member of the British Computer Society, Paul has lectured extensively to customer, technical and user groups as well as inside ICL.

Glossary of Terms

ADPCM – Adaptive Digital Pulse Code Modulation
AM – Amplitude Modulation
CDMA – Code Division Multiple Access
CEC – Commission of the European Communities
CEPT – Conference of the European Postal and Telecommunications administrations
CTO-3 – Cordless Technology 0, 1, 2, 3
DCS1800 – Digital Cordless Technology 1800 Mhz
DECT – Digital Cordless European Telecommunications
ETACS – Extended Telecommunications Analogue Cellular System
ETSI – European Telecommunications Standards Institute
FDMA – Frequency Division Multiple Access
FEC – Forward Error Correction
FM – Frequency Modulation
FSK – Frequency Shift Keying
GSM – Groupe Speciale Mobile
ISDN – Integrated Services Digital Network
ISM – Industrial Scientific and Medical
ITU – International Telecommunications Union
LAN – Local Area Network
PBX – Private Branch Exchange
PCN – Personal Communications Network

PM – Phase Modulation
PMR – Public Mobile Radio
PRBS – Pseudo Random Binary Sequence
PSK – Phase Shift Keying
SST – Spread Spectrum Technology
TACS – Telecommunications Analogue Cellular System
TDM – Time Division Multiplexing
TDMA – Time Division Multiple Access
VHF – Very High Frequency
WAN – Wide Area Network
WARC – World Administrative Radio Conference

Communications Technology for the Retail Environment

Edwin Turner
ICL Retail Systems – Bracknell, UK.

Abstract

The retail environment is very familiar to most people, but few are aware of the technology which supports the store operation beyond the cash registers and Electronic Point of Sale (EPOS) systems. Fewer still appreciate the amount of information processed and acted upon following the simple act of scanning a can of baked beans at the checkout of the local Sainsburys or Tesco.

This paper looks at the environment in which in-store IT equipment operates and discusses some of the factors limiting its further deployment. Significant among them is the difficulty of interconnecting IT equipment in this hostile environment. The paper goes on to describe how cordless communication technology is poised to take on this demanding task.

1 Introduction

Since the early 1980s, retailers have been seeking to improve their business profitability by the application of technology. They first sought to improve the efficiency of the selling cycle by the introduction of EPOS equipment to de-skill the selling process. Retailers soon realised that as a by-product of this equipment, large amounts of accurate information were available on the types and amount of goods sold and the customers who bought them. They began to develop increasingly complex systems to manage the flow of such goods throughout the cycle from procurement to customer delivery (Pickworth, 1991). Such developments occurred alongside the emergence of supermarkets and superstores, where the cost of the technology could be justified by the increased turnover of individual stores.

More recently, as the costs of equipment have fallen, smaller outlets have been able to afford the benefits of EPOS. The competitive advantage generated by tight control of stock and of the selling process has diminished; retailers are now looking to extend the use of technology to other areas to regain that advantage. Increases in computing power have been perceived as "free", but the ability to exploit this power fully has been constrained by the problems of interconnecting systems to make optimum use of the data. Within the office, the LAN has become a major strategic component of the IT infrastructure. Complex cabling systems interconnecting office data-processing equipment have been developed alongside networking components. It has been equally clear, however, that solutions offered in the office market are frequently inapplicable in other environments such as retailing.

2 The Retail Environment

A retail environment has a number of characteristics which differ markedly from that of the typical office, being often closer to that of a heavy industrial factory. These characteristics have been evaluated in several surveys, and can be summarised as follows:

2.1 Retail outlets use widely dispersed connections in an extremely hostile electromagnetic environment. The connection density of equipment is such that any form of saturation cabling similar to that in an office (Flatman, 1988) is both impractical and prohibitivly expensive. High switching loads, such as the bakery ovens and refrigeration compressors present in large modern supermarkets, generate high levels of electromagnetic interference. Static electricity is also a hazard in the dry atmosphere of a modern store, often fitted with nylon-mix carpets. Such interference is easily picked up on any connection systems not designed to cope.

2.2 Retail areas are visited by the general public, and high standards of safety are demanded for their protection. Particular care must be taken over both electrical and physical installation, bearing in mind that unsupervised members of the public have legitimate access to most areas of the store. Injuries of any sort to customers can have a serious effect on the reputation of a retailer and consequently on his business profitability.

2.3 The visible end-product of the selling process is the bill and customers are extremely quick to notice any discrepancies in the amount charged. No-one likes to be overcharged (it appears that more mistakes result in overcharging than vice versa!). Consumer protection legislation is strongly biassed towards the customer and, once again, no retailer can afford the effects of poor publicity on their reputation. Consequently data integrity is of prime importance.

2.4 Many retail outlets, particularly for food, are now totally dependent on EPOS systems to transact business with customers. This is most obvious in the large supermarket/hypermarket environment where it is now almost impossible to trade if there is a failure of the EPOS system. Since both equipment and installation require a very high level of system reliability, most systems now use levels of redundancy and fault tolerance better than most of the DP industry. Redundancy extends from EPOS hardware, through cabling infrastructures to supply and distribution of electric power. It should be noted also that the users of such equipment are not highly qualified as DP operators.

2.5 Retailers are expert in cost negotiation with suppliers, be these of garments, tinned fruit or EPOS systems. Few electronic support systems contribute to the primary objective of pursuading the customer to purchase goods, and so are effectively "grudge buys". As such the retailer ensures that he pays as little as possible for them. Systems will generally be written off over a period of at least seven years, while it is not uncommon to see store cabling infrastructures which have been in place for 20 years.

2.6 As shop opening hours extend there is a tendency for the time available to change the store infrastructure to diminish. The recent move to seven-day trading in UK supermarkets is not unique, stores in the USA have been so trading for many years. It is reasonable to expect UK and other European stores to open for 22 hrs per day in the future, as occurs now in the USA. This long period of trading increases the amount of data to be processed while at the same time reducing the time available for processing and communication with central office or for changing store layout. It should be noted that most retail operations are controlled largely from the centre of the operation, with relatively little local autonomy (Pickworth, 1991).

2.7 Shopping malls have added to the requirements of the EPOS system. Several stores within the complex will often belong to one group, particularly in fashion and jewellery areas. It is more cost-effective and much easier to obtain benefit from EPOS, if members of the group could behave as if all were connected as one single system. It is often difficult to arrange physical interconnection of the various outlets because of the structure of the complex. It is also difficult because of the transitory nature of the outlets, some operate in one place for as little as a year before being closed, refitted and reopened as another store within the group.

3 Requirements of the Retail Process

As Pickworth (1991) showed, the retail process involves far more than the simple exchange of goods for cash. It includes areas such as security, all forms of environmental control, the control and delivery of stock and also promotion and advertising. The following looks at each of these aspects and tries to assess where new methods of communication could improve operations or assist in attracting new customers. A diagram incorporating elements of both cabled and radio connection is shown in Figure 1. Some of the components are discussed below.

3.1 EPOS Facilities

This area of store operation has already benefitted significantly from technology. Barcode scanning has improved the speed of checkout to the extent

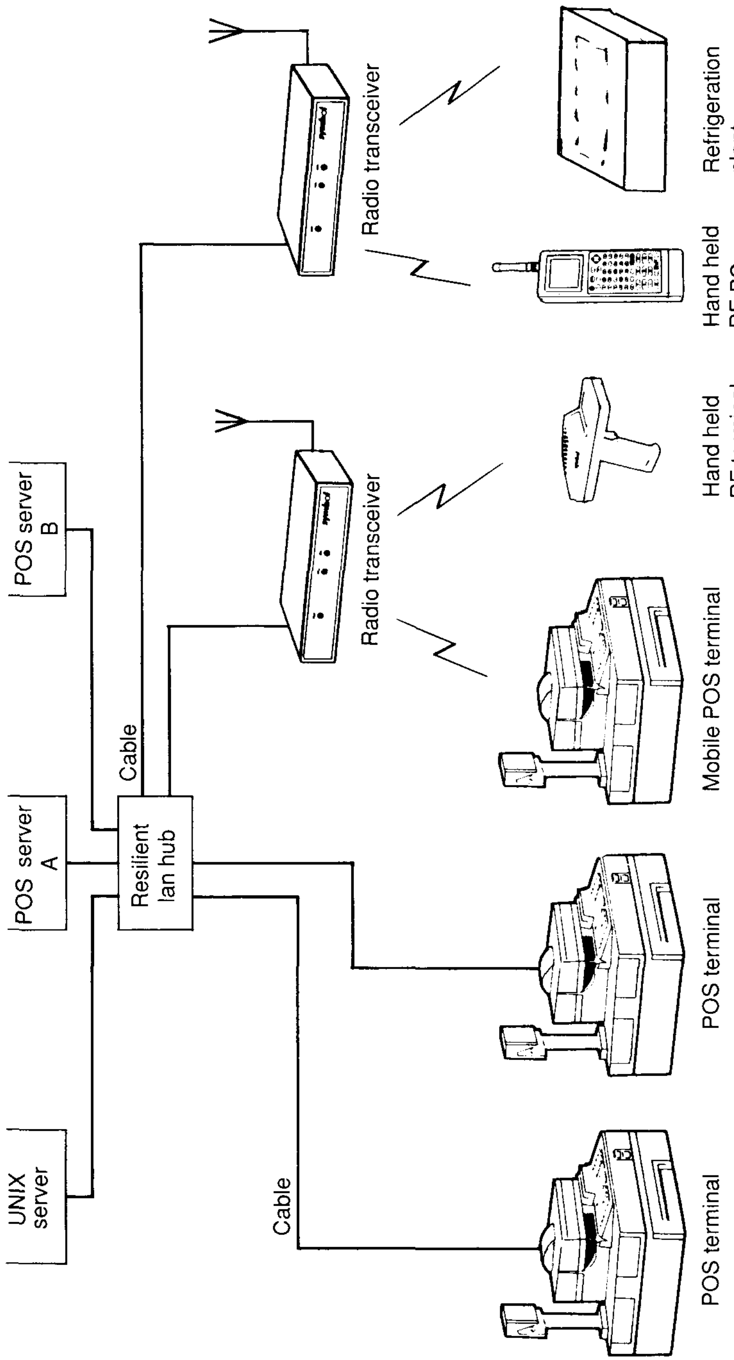


Fig. 1 Future Store Interconnection.

that the limiting factor now appears to be the time taken to open a carrier bag and pack the goods. A good operator can scan items at the rate of almost one per second, while the automation of price entry means that transaction accuracy is very high. Hand-held scanners offer similar benefits to a non-food store where it is often impractical to scan items passed over a flatbed unit, this being particularly true in DIY supermarkets. Current EPOS systems are mainly fixed in layout and location, limited by the large size and power requirements for the checkout peripherals, and the need for a cabled connection to the in-store computer system.

Two recent technologies offering the possibility of significant change in this area are portable (laptop or notebook) computers and radio connection. It is now practical to build a mobile checkout connected to the main system over a radio link. This offers a retailer the opportunity to rearrange the store layout more frequently, suiting the EPOS sites to the store layout rather than vice versa. It also allows flexibility, in that equipment can be added at times of peak loading, e.g. during sales periods or at seaside locations in the summer, while releasing valuable floor space for display purposes at other times.

Cableless connection allows the rapid deployment of EPOS equipment in areas where it would previously have been impractical, e.g. at a promotion in the store or even in the car park. If the standards of radio system chosen are compatible with those of the PSTN, it becomes practical to connect EPOS equipment at an exhibition back to a remote central system to allow further processing of its data.

3.2 Store Security

Security in a retail store covers a number of areas, all of which can benefit from new communications media. Many stores are protected by CCTV systems, which can be monitored locally or remotely via wideband communications, such as ISDN, allowing better coverage by increasing the number of cameras while reducing the number of monitoring staff needed.

Theft from the store during normal trading is an escalating problem, and the trend towards self-selection of goods coupled with reductions in staffing levels exacerbate the situation. Techniques are available for electronic tagging of items which are now used for high value items such as fashions, electronic and camera equipment. Such systems will be linked in some way to the store communications infrastructure in order that appropriate security CCTV cameras or staff are made aware of attempted theft and can take appropriate action. The cost of such electronic countermeasures will drop to levels where they become commonplace on lower value items. However, it will probably take a radical change of approach to secure all the items in, say, a supermarket.

3.3 Environmental Control

The modern store is adding complex systems to monitor its various environments both in order to reduce energy costs and to ensure that goods are maintained in saleable condition. This latter aspect is particularly important in food retailing, where new food preparation, handling and storage legislation seems to appear from government bodies (particularly the EC) almost daily. Customer expectations concerning the quality and presentation of items have increased markedly in recent years as stores compete for custom.

New monitoring systems tend to be manual or self-contained systems. Stores look to integrate automatic monitoring systems, but the rate at which new requirements are added means that it is difficult to connect these systems by cable, since the cabling infrastructure would be changing continually. This would also therefore appear to be an ideal application for radio communication, particularly since the data throughput required for such monitoring will in general be low, and low cost technology can be used.

In the food sector of retailing, monitoring requirements now extend further along the supply chain; it will soon be necessary to monitor and trace all perishable food from preparation, through supply and distribution, to its display and sale. This calls for monitoring on transport facilities (lorries etc.). If such monitoring equipment is to be on-line radio communication is the only practical method. Given that approximately 50% of the goods in a large supermarket are perishable, the size of the problem can be appreciated.

3.4 Product Pricing and Labelling

New consumer legislation coupled with the damaging effects of bad publicity over marking of prices have stimulated keen interest among retailers in electronic presentation of price and other information. Very low cost, ease and mobility of installation, low power and high legibility provide a severe challenge for manufacturers of electronic labels, though the potential market is immense. Consider a large supermarket carrying over 20,000 different items, each requiring a label. Multiply that by the number of outlets in that supermarket chain, and the scope of the opportunity becomes apparent. Technology exists to provide the label itself using an LCD display and CMOS microcontroller, but as yet the interconnect medium has proved a stumbling block. Cable-based interconnection schemes have been tested but were found too expensive and inflexible. Several schemes are under development which connect each label or small group of labels via a low performance radio link. None has yet proved capable of meeting the cost/performance target of less than £8 for a label with two-way communication.

3.5 *Stock Control*

While EPOS systems have provided data on products sold, other factors need monitoring to provide an accurate indication of stock. These include the effects of pilfering and accidental damage, and the status of date-stamped items, particularly perishable goods. Most of this information is only available from the store floor where mobile, hand-held terminals are used, either custom designed or portable PC-based and capable of scanning barcodes. Output from these terminals is loaded into the main EPOS system as a secondary operation, which has a number of flaws, most notably lack of access to the main store database. It would be preferable for example to allow an operator on the floor to mark down items approaching their sell by date, and to reflect this immediately in a price change at the till. This requires an interaction between the mobile terminal, marking down the price on the store floor, and the in-store computer controlling the price files in the EPOS terminals. Radio communication for the mobile terminal is the only practical method for this communication.

Terminals in the loading area used by personnel monitoring goods being delivered also need to communicate both with local computing and with remote distribution facilities; if such terminals are mobile greater efficiency is possible. It is normal practice in stores to use the main store computer as the gateway to all out-of-store communications, using duplicated dial-up or leased lines to head office mainframe computers.

3.6 *Customer Support Terminals*

There is an increasing trend to use computing power in the store to provide services for the customer. Both the assistant and the customer may interact directly with terminal equipment so as to choose or request information on products or services. This form of interaction calls for high levels of reliability and ease of use, and powerful computers are used to support these customer interfaces. These terminals need to be connected to or via the main store computer; while in many cases they can be connected by cable where it is already available or easily installed, it is often more cost-effective to install isolated terminals using radio. This also applies to isolated EPOS equipment in e.g. the wine store or the delicatessen.

Multimedia technology involving still and moving image, data, voice and music is now being introduced onto the store floor to assist selling. The first implementations are found in fashion and soft furnishing areas. Computing power coupled with video technology provides "what if" scenarios to assist the customer to choose products and services. As an example a customer might ask: "what would my room look like with wallpaper of this design, curtains of that and furniture coverings of a third?. If it looks nice, are all the items in stock in sizes to suit my rooms and what would be the cost? Finally could I place the order, pay by credit card no. xxxx and have the goods delivered and installed on Tuesday next?" Customers expect pictures of goods or installations with quality equal to, or better than that of a normal TV. Such a multimedia terminal requires video, usually a remote controlled videodisk, and access to a local store computer for stock position, installation and transport organisation. Through the local store computer it also needs access to a remote computer for the stock position at a distribution warehouse, and also for credit card verification. All connections within the store can be made over radio, including the control of the videodisk, making for a multimedia service-point requiring only mains power.

3.7 Wide Area Communications

Many store chains cover a wide geographical area and most are controlled largely from head office so that efficient, wide-area communication is essential. Currently major links use dial-up PTT connections or leased lines, always duplicated for resilience. Introduction of higher performance PTT services such as ISDN can offer lower connection costs. Small numbers of EPOS terminals in speciality stores could be linked directly to controllers at head office using ISDN links, thus minimising purchase and support costs for in-store controllers. This will not happen until the network operators can demonstrate a highly reliable ISDN service, since the stores ability to trade will depend on the service, and retailers are cautious by nature.

With the development of low-cost satellite systems, some more adventurous retailers are exploring the potential benefits of their use. Much data sent from head office to store is common to all in the chain; an example might be a video-based catalogue for multi-media systems. Using a satellite would allow broadcast transmission to all stores simultaneously, using control information within the broadcast or other wide-area connections to provide selective customisation. Its primary advantage is ease of installation, there being no need to wait for expensive leased lines which may indeed not be available. The concept is already in use in one major sector, bookmaking, where it has proved extremely successful. Its future in other sectors could be equally exciting.

4 Current Interconnection Infrastructures

Current store cabling infrastructures have developed over a number of years in a seemingly haphazard manner. Cabling of EPOS systems has to date been proprietary, with multidrop LANs using a wide variety of cable types, usually screened and very rugged. Voice, whether telephony or intercom, has used one or two separate networks based on telephone cabling practice. Security has operated over a third network, usually coaxial cable. Finally monitoring equipment will be connected over yet other networks using cable of all types from simple unscreened twisted pairs to fibre optics. Cabling will be located either in false ceilings or in ducting built into the solid floor.

Cabling located in the ceiling void has the advantage that there is adequate space to add further cables if the store equipment or layout changes, but has a number of disadvantages. The void sometimes contains asbestos which if disturbed will need to be removed. This is particularly true of department stores built in the 1950s/60s; the cost of its removal is very high, far exceeding the cost of cabling. There is the problem of linking equipment to cables in the void, neccessitating unsightly poles from equipment to the ceiling. Finally there is the ever-present danger of disturbing cabling already there with a consequent malfunction of some other equipment in the store.

Cabling in floor-ducting does not have these problems but faces others. Ducts are rarely in the correct place, and store layout will generally be determined by duct position rather than vice versa. The ducts will often be too small to carry the increasing number of cables required, and the cables themselves may be at risk due to the effects of cleaning fluids used on the floors which get into the ducts. The cost in terms of disruption to the store of adding new ducts can be very large if the store has concrete floors, as is common in large food supermarkets.

Finally, there may be instances where neither of these solutions is practical. One example could be a listed building where there are restrictions on changes to the building fabric. The authorities would not permit damage to panelled walls or elaborate plaster ceilings due to installation of cables. If floors are carpeted, there appears little scope for interconnecting EPOS or similar equipment at all without unsightly cable trunking.

Thus, cableless communication for IT systems can offer significant benefit in the retail environment. Many large retailers are experimenting with connection systems based on radio and infra-red transmission. The latter has had little success due to the difficulty of installation and the maintenance of line-of-site to ensure reliable communication, even though it does not require regulatory approval. Radio links do not suffer from this line-of-sight drawback but to date have been severely hampered by lack of approved frequency spectrum. With easier availability of equipment and spectrum, interest in the use of radio has increased for all forms of in-store interconnection.

5 New Methods of Interconnection

As noted, while infra-red technology has been tried for different purposes, achieving adequate range for mobile application without the use of high-powered sources has hampered its use. Such sources tend to be based on lasers rather than LEDs and a perception that "lasers are dangerous" has not helped, though the laser power used in these applications (Ref IEC 825) is quite safe.

Radio techniques for use in data communication have largely come from two different industries, and each has approached the problem with a different set of priorities.

5.1 High performance solutions have been developed largely by companies in the military supply industry. An effect of the lessening of world hostilities has been a fall in demand from their traditional customers. They have moved into the commercial market as a method of diversification. They have concentrated largely on the performance, reliability and data integrity of their products at the expense of cost and time-to-market. The products use state of the art "spread spectrum" technology operating presently in the 900 MHz and 2.4 GHz ISM band. The link has a range in excess of 200 metres in broadcast mode with data rates up to 2 Mb/sec and data integrity comparable with that of a cabled LAN.

Though they are falling, costs for such units currently exceed £1000 per connection. The products are targeted at LAN replacement, and the power needed to meet this performance rules them out for mobile use. Nevertheless their integrity and range have persuaded a number of retailers to try out their use for EPOS connection and one large retail chain is installing them for non-food checkouts. It should be noted that apart from meeting type approval for factors such as power output and frequency usage, there are as yet no standards governing these products and interworking between different manufacturer's products is not possible.

Further bands are presently being negotiated in the 5.2 GHz, 17 GHz and 60 GHz regions, which will give correspondingly higher performance but lower range.

5.2 Lower performance products are emerging from the telecommunications industry as it seeks to exploit its massive investment mobile phone technology. Their design philosophy is almost the opposite of that of the military suppliers, placing emphasis on low cost, low power and adherence to transmission and protocol standards at the expense of range and performance.

5.2.1 "First generation" products now emerging use the CT2 voice telephony standard based on the 900MHz voice band. Range is limited to approximately 50 metres, requiring use of a cellular network of base stations. The system is less suited to mobile use, since moving from one cell to an adjacent one loses the connection. Data rate is limited to 32kb/sec. and error rates do not meet the data integrity requirements for EPOS type connections. Nevertheless their low cost, estimated at under £200 per connection means that they would be suitable for low rate monitoring applications. They would use a "dial-up" mode similar to normal voice telephony. This can be an advantage where a large number of connections are used each only infrequently, allowing maximum use of the shared spectrum. Other benefits include the ability to use the same base station to carry voice traffic for in-store communication, and to connect to the wide area network for remote access.

5.2.2 The "second generation" products just emerging from the development labs meet the new DECT (Digital European Cordless Telephony) standard in the 1.8GHz voice band. This system has been designed to incorporate data transmission as an integral part of the standard. It employs 32 kb/sec "blocks" of bandwidth, up to 12 of which can be concatenated to give performance on demand ranging from 32 kb/sec to 384 kb/sec per channel. Range is up to 50 metres, thereby still requiring a cellular base station layout as with CT2, but the protocols used give an error rate comparable with Spread Spectrum. Cost should be comparable with CT2, though, since the latter is further ahead in its life cycle, DECT will probably remain more expensive than CT2. Voice can operate across the same base station infrastructure using appropriate handsets.

A significant advantage of DECT over CT2 is the ability to "roam" between base stations without losing the call connection, making it more suitable for *mobile* applications such as stock checking as opposed to *movable* applications such as monitoring where the unit is placed in position, the connection established and the connection closed before the equipment is moved.

5.3 Other systems are available for lower cost uses, but most have problems operating in electrically noisy environments, and, using lower modulation frequencies, are more susceptible to interference, whether accidental or deliberate.

6 Conclusion

Further improvements in productivity in the retail industry will require increasing use of IT power, but deployment of such power will be severely restricted without improvements in system interconnection. The environmental restrictions, the need for maximum flexibility, speed and ease of installation and a need for a mobile processing capability indicate that cableless connection would appear to offer the best solution to these requirements. Further deregulation of the telecommunications industry together with the development of lower cost radio technology indicates that radio based communications will be a cost effective way to augment current cabling technology, although it will be some time before it completely replaces cable. Most leading retailers are evaluating different forms of radio technology for in-store use and some are actively deploying it in new installations. Growth will continue slowly until standards for radio based protocols evolve, at which time lowering component cost and easier interworking with cable-based LANs will speed up its introduction rate dramatically until eventually it becomes the dominant communication medium.

References

FLATMAN, A.V. Universal communications cabling – a building utility, *ICL Tech. J.* Vol. 6 No. 1 pp. 117–136, 1988.

PICKWORTH, I. Experience of managing data flows in distributed computing in retail businesses, *ICL Tech. J.* Vol. 7 No. 4 pp. 718–731, 1991.
IEC Standard 825 deals with the radiation safety of laser products.

Biography

Dr. Edwin Turner

Ed Turner holds a B.Sc in Computer Science and Ph.D in Computer Engineering, both from the University of Manchester. He joined ICL in 1973, working on the evaluation of mass storage media for ICL's new range of mainframes at Kidsgrove.

Since moving to Bracknell in 1978 he has undertaken development roles on a variety of products ranging from the System 10 and System 25 minicomputers to the One-per-Desk workstation, as well as working in a strategic role in a number of areas. He joined Retail Systems in 1990 where he has responsibility for their interconnection strategy as well as undertaking a number of related development roles.

RIBA – A Support Environment for Distributed Processing

L. E. Crockford
ICL Product Operations, Stevenage, UK
A. Drahota
Government & Major Companies Division, ICL, Slough, UK

Abstract

The computing needs of large and medium size organisations today exceed the capabilities of individual computer systems and division of processing between multiple systems is necessary. Most organisations are distributed geographically and computer users within them expect systems which reflect the distributed nature of the human organisation. It is rarely possible to separate the data and processing without any inter-dependencies between systems. Thus inter-working is needed between applications running in separate host computers.

Distribution of processing introduces many complexities for system designers, developers and managers. The RIBA support environment is based on the ANSA architecture and tackles a wide range of issues. It aims to support distribution of processing and data in heterogeneous environments of host computers, database management systems, programming languages and communications systems.

Due to the wide scope of RIBA, this paper provides only an overview of the approaches adopted, in particular transparency mechanisms and language tools. It then concentrates on data transparency, an RIBA development which conceals from the developer the logical structure of information implemented in various databases, and which enables applications to be developed against conceptual models of information. The aim is to de-couple an application program from any particular type of database, thus enabling future migration and safeguarding portability of the application.

Collaborative development of RIBA is under way with user organisations to demonstrate the relevance and effectiveness of the RIBA approach in actual business situations.

Introduction

RIBA is a prototype implementation of a software environment developed within two Esprit II projects (RICHE* and BANK '92**) during 1989–1991 [DRAH90]. The two projects concentrated respectively on the broad requirements of health care and financial markets. The name (RIBA) is derived from the two project names and as such does not carry any meaning.

RIBA addresses the key requirements identified from users, developers and integrators of systems in the two industries. These requirements were subsequently verified with a range of organisations from other industries and were universally confirmed.

1 Requirements Addressed by RIBA

Information Technology (IT) is evolving rapidly and the pace of development is causing considerable difficulties to its users. Individual application systems often require significant resources (time, expertise and money) to develop and this investment needs to be protected. At the same time, organisations wish or require to take full advantage of new technologies, for example for economic or competitive reasons. Portability of application systems is an important requirement.

Most large organisations are distributed geographically, and consist of departments carrying out a variety of tasks. Increasingly, users of IT expect systems which reflect the distributed nature of the human organisation. The ability to distribute processing and manage such distributed systems is a key requirement.

Today, the processing needs of large and medium size organisations are exceeding the capabilities of individual computer systems. Therefore, division of processing between multiple systems is necessary. It is rarely possible to separate the data and/or processing without any inter-dependencies between sub-systems. Therefore, inter-working is needed between application systems running in separate host computers.

There are varieties of computer hardware and software products available in the market place. Different systems are better suited to particular tasks and users wish to acquire the best systems for their processing needs. Therefore, the ability to integrate and support, easily and cost-effectively, heterogeneous environments of computers, operating systems, database man-

*ESPRIT Project 2221, Reseau d'Information et de Communication Hospitalier Europeen.
**ESPRIT Project 2476.

agement systems, communications systems and application packages is extremely important.

Many new application systems address complex user requirements, for example complete business processes rather than just individual tasks within them. Such new systems are often technically complex and rely on the integration of older systems with new applications and technologies. Tools are required to simplify and speed up development of new code, as well as to assist in the integration task. Such tools must hide much of the technical complexity of the system from the developer, so that he or she can concentrate on the logic of the system without concern about subsequent implementation options. Only thus will it be possible to design and develop systems which are readily portable between, possibly quite different, technologies. Furthermore, it is essential that the resulting systems are simple to maintain and support.

Very many organisations are extremely dependent on their IT infrastructure to carry on their normal business. In some organisations the dependency is such that even relatively minor disruptions in service can be very damaging in terms of profit or costs, customer satisfaction or indeed human safety. Therefore, systems are required which can tolerate faults, which are able to recover rapidly from failures and which can be enhanced or re-configured whilst application services continue to operate. At the same time, such flexibility is required at a reasonable cost, for example only a few organisations are willing and able to fund complete duplication of their computer installations.

Finally, the scale and complex nature of many of today's computer systems place major strain on their user. The costs of system maintenance and support are beginning to use up all available resources; research shows that many large organisations are expending circa 90% of their IT budget on keeping their existing systems running and effective in the changing business environment. This burden is one of the causes of the existing huge back-log of application development. Improving systems maintainability is a key requirement.

None of the above-mentioned requirements is new, but it is useful to consider them as the backdrop against which RIBA is being developed.

2 RIBA Approach

RIBA tackles the identified requirements in several ways. Firstly, it offers an architecture for construction of distributed systems. This in itself leads to a number of benefits, for example by providing a standard for application inter-working which can prevent proliferation of different inter-working schemes being adopted by system integrators. The RIBA architecture is based on the Advanced Networked Systems Architecture (ANSA) [ANSA89] which is itself driving the International Standards Organisation's (ISO) standardisation programme – Open Distributed Processing (ODP) [ISO89].

The characteristics of the RIBA architecture can be summarised as follows:

- based on the client-server model of computing
- systems can be organised into many federated domains
- complexities of distribution are concealed from users and developers by transparencies [ANSA89]
- *location transparency* allows client programs to be developed without knowledge of the locations of their servers
- *access transparency* ensures that the interactions between client and server programs remain identical, semantically and syntactically, regardless of the communication route between them
- *physical data representation transparency* ensures that data are received in their local representation and that applications need not perform data conversions dependent on the differences between the client's and server's hosts' characteristics
- *logical data representation transparency* allows applications to be developed to the conceptual model of data (e.g. entity-relationship-attribute model) and thus insulates the applications from the logical representation of the data (e.g. relational tables, Codasyl sets, image files, etc.).
- *migration transparency* allows applications or data to be relocated within the networked environment without any need to modify their clients
- *replication transparency* will enable multiple servers to be run without their clients needing to be aware of such replication. All coordination of the replicas will be managed by the transparency mechanism

2.1 Transparencies

In general terms, a transparency mechanism can be defined [RIBA92] as that which allows the client of a server to:–

have a model of the service provided by the server which, to some extent, is different from reality

invoke the service and anticipate responses from the server in terms of the client's model

It is the usual case that the client's model of a service is some abstraction (simplification) of reality and thus transparency mechanisms have the role of concealing complexity. An essential ingredient of a transparency mechanism is a means for transforming between models, this is known as a *transform.*

The most important transparencies are concerned with hiding the effects of distribution, of heterogeneity and of potential changes. Examples include masking of the differences between, say, Codasyl and Relational databases, or enabling applications to be written without any knowledge where other services they will require exist in the network.

2.2 *Models and Transforms*

A *model* is a description of a state or of a state of processes, using a set of concepts and expressed in some form of language.

A *transform* is an algorithm or a collection of rules that when applied to a model will produce another model that differs from the original in some respects e.g. a change of concept, different language or an abstraction. If a transform is reversible its inverse can be applied to the new model in order to regenerate the original model.

Figure 1 below shows the general scheme of a transparency mechanism.

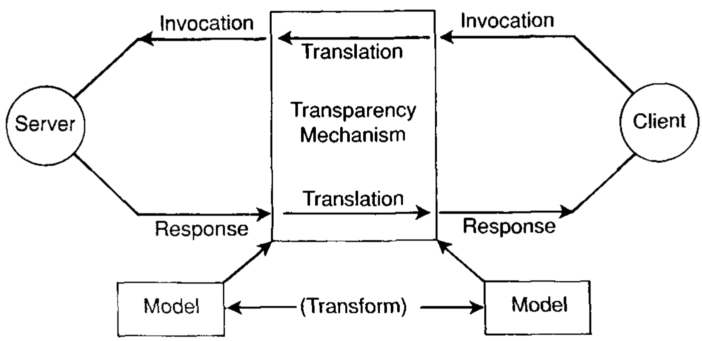


Fig. 1 Transparency Mechanism

A model of the service, held by the server, is supposed to exist, together with some reversible transform that produces from it a different, usually simpler, model held by the client. The transparency mechanism has access to both models and to the rules or algorithms that define the transform and its reverse i.e. provide the ability to translate between the models from either direction.

In operation the mechanism receives invocations from the client and applies the reverse transform in translating the invocations into the context of the server's model. Similarly, responses from the server may be translated, using the transform, into terms understood by the client.

2.3 *Application of Transformers*

The basic scheme can be applied through three different techniques, either during the development epoch or at run-time.

Compilation is the most useful technique where models of services are reasonably static. A client program is originally constructed in accord with its (simple) model. A compiler or pre-processor performs all the translations of clients' invocations and response expectations and reconstructs the client so that it will operate in the context of the servers (complex) models.

Compilation is the most efficient of the techniques but it suffers from the disadvantage that should any service change its model then all of its client programs must be reconstructed by compilation.

The second technique is known as *binding*. This is effective where the differences between a service model and the clients' model of the service are slight.

It operates only at the time of establishing a binding between the client and the server. At this time the model used by the client is modified by the transparency mechanism so that it equates with the server's model. Subsequent interactions between them are thus in the context of the shared model. An example of this approach is the RIBA Trader which modifies models held by clients so that they become aware of the location of the server.

Finally, there is *dynamic translation*. Here, the transparency mechanism is permanently inserted between server and client. All invocations from clients are translated before being passed on to the server and responses are similarly treated.

This use of what in software construction is variously called transformer, agent, filter, protocol converter or relay carries a penalty of a performance overhead on each invocation. However, this is a trade-off for the long term immunity from change i.e. the client's model remains constant so long as there is a valid transform for the changing service model. An example of this type of transparency mechanism is the Remote Procedure Call (RPC) service, which enables programs to invoke each other by means of a simple procedure call whether the programs are co-located within the same host computer or separated by some kind of a network. The RPC mechanism translates any calls between programs at run-time onto the most appropriate available transport service and ensures that a consistent appearance of the whole system is maintained for its clients.

In the RIBA implementation, all three types of transparency mechanism are exploited.

2.4 Key Assumptions

Development of application systems targeted to run in distributed environments is very different from that for centralised systems. Analysis shows, that many, indeed most, of the assumptions traditionally made by developers of centralised systems become invalid when distribution is introduced. For example, in the case of failure, the designer cannot assume that all parts of the system will have failed at the time of the fault, because in a distributed system many of the distributed components may continue processing while one is out of action. This significantly complicates recovery procedures. Other examples include assumptions about homogeneity of names and file representations, about use of shared memory (difficult when programs are separated by a network and hosted on architecturally different computers), and about processing occurring serially rather than, possibly, concurrently. More extensive discussion about the differences between centralised and distributed approaches can be found in literature about ANSA e.g. [MARS91].

In RIBA, it is assumed that systems will not only be distributed (a centralised system then becomes one extreme example of distribution), but also that the host computers and their operating regimes may be different. It is also assumed that large distributed systems may consist of many domains, themselves distributed, which may adopt different data administration and system administration schemes. Thus, RIBA does not insist on an organisation-wide standard for systems administration, while it supports inter-working between such different domains. Allowing this type of heterogeneity is particularly important for organisations requiring an evolutionary approach to systems development or needing the ability to integrate rapidly different systems, as may be the case in merger and take-over situations. The technique used is known as *federation*, and it again originates from ANSA.

Concealment of many of the physical characteristics of the host environment also means that designers are unlikely to "bend" the system design towards any particular technology and hence that the design has a much better chance of surviving when the underlying technologies become obsolete.

The provision of transparencies allows systems to be re-configured without any impact on the application code, for example services or data can be relocated without impact on their clients. Application clients can be developed on RIBA which automatically adapt to re-configuration of their servers, indeed most services can continue running whilst reconfiguration is taking place.

Transparencies obviate development of conversion code by application developers who are no longer concerned with the origins of information. The latest developments in RIBA are beginning to de-couple the application code from the logical structure of the data and thus allow data restructuring and migration between different databases, again without impact on the application code. These developments are the main subject of this paper and are described in section 3.

2.5 *Supporting Tools*

A number of development tools are provided in RIBA which automatically generate code from high-level statements supplied by the developer. The developer is supplied with two languages for expressing the requirements and characteristics of programs.

Interface Definition Language (IDL) is a programming language independent notation for describing interfaces provided by servers and required by clients. IDL definitions can be compiled to generate all the code necessary for clients to communicate with servers.

Distributed Processing Language (DPL) is an extension to conventional programming languages and it allows the developer to express inter-working requirements within the client and server programs. DPL statements are typically used for forming bindings between clients and servers at run-time.

Thus RIBA eliminates some coding effort, particularly in areas of complex system code.

2.6 *Compatibility*

All the RIBA mechanisms are optional, and the platform and application systems running on it can coexist with existing systems in their host computers. Existing applications can be integrated with RIBA applications either by means of encapsulation, whereby they are turned into RIBA objects, or through "transformers", which are RIBA objects offering a conventional appearance to the external applications.

The next section concentrates on the RIBA Information Service, which is the part of the system which conceals much of the complexity normally encountered when programming with distributed data that is managed by different database management systems.

2.7 *RIBA Demonstrator*

A demonstration system was constructed at the end of the Bank '92 project to show some of the capabilities which RIBA offers.

The demonstration system shows a set of simple banking applications running in a distributed environment consisting of some 15 computers (from ICL, Sun and Next) running 6 different operating systems. The demonstration also includes 3 different database management systems (Ingres, Oracle, Sybase) as well as an image database (ICL Powervision) integrated into a virtual database.

The demonstration shows inter-working, re-configurability and application independence from many characteristics of the host environment.

3 The RIBA Information Service (RIS)

The RIBA Information Service permits its clients to view all of the information managed by the service as though it were contained in a single virtual and co-located database. This is a considerable simplification of the traditional approach where the developer of an application program is required to take cognisance of any separation and dispersal of the information to be manipulated.

Application programs that use the service are designed and constructed solely in terms of an Entity Attribute Relationship (EAR) model of the information to be accessed and manipulated. This again is a simplification as designers are not concerned with the structures that are employed to hold their information and use a single, simple language for its access and manipulation. These applications enjoy a long term immunity from the effects of re-organisation and re-configuration, needing to be maintained only when significant changes are made to the models of the enterprise and its information.

The RIS uses the basic mechanisms of the RIBA platform for location, access and physical representation transparencies with respect to data i.e. the interface trader and remote procedure call. In the current implementation of the RIS the logical data representation transparency is provided by dynamic translation of messages.

Figure 2 shows the service in conceptual outline.

Standard, commercially available, systems are used for data management such as Ingres, Oracle, Sybase, PowerVision, IDMS etc.

The RIS provides each of these systems with front end processors known as *Logical Information Servers* (LIS). These servers present a RIBA conformant interface into the database management system and also perform some message translation that varies in degree with the type of data management system.

The totality of the data may be separated into discrete *partitions* for reasons of security and privacy.

A server, called the *Conceptual Information Server* (CIS) is a client of a group of logical servers for a data partition. It offers an interface to application clients and will accept messages from those that are authorised to access the data held in the partition. These messages are translated by the CIS and projected, in a co-ordinated manner, over the logical servers. Responses from

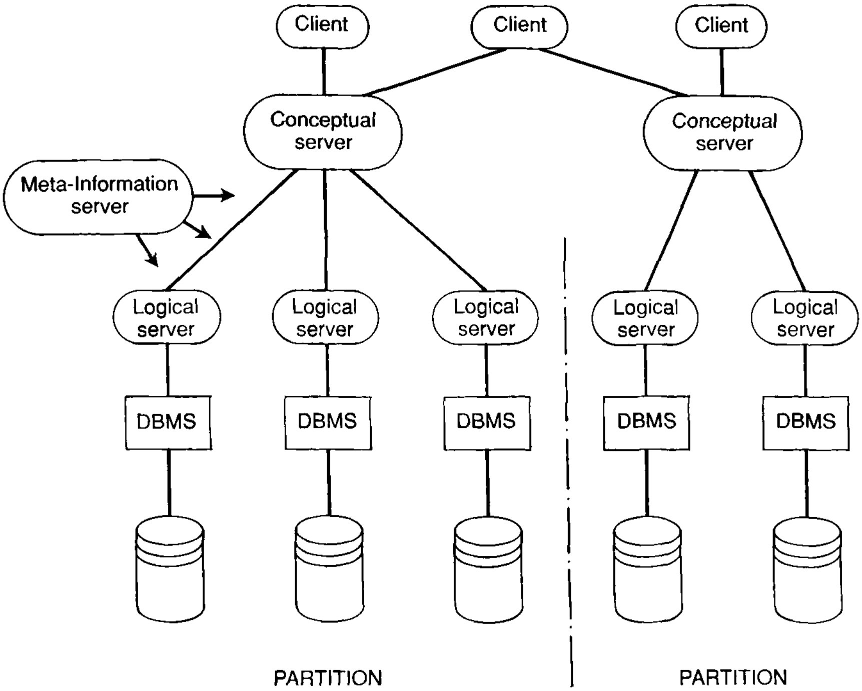


Fig. 2 RIS outline

the logical servers are then collected by the CIS and composed into a unified response to the client.

A *Meta Information Server* (MIS) manages the models of the information and of its separation and distribution and supplies them, on demand, to the other components of the information service.

The next section describes these models and gives an outline of the modelling language.

3.1 *Information Models and Mappings*

The model of information and its separation is layered thus:

1 Conceptual models
  Conceptual to Logical mapping
2 Logical models
  Logical to Database Schema mapping
3 Database Schemata

In layer 1 there is a model, at the conceptual level, for each of the separate partitions of information. In layers 2 and 3 there are logical models and corresponding schemata for each of the underlying databases.

Figure 3 shows how these models are used by the components of the service.

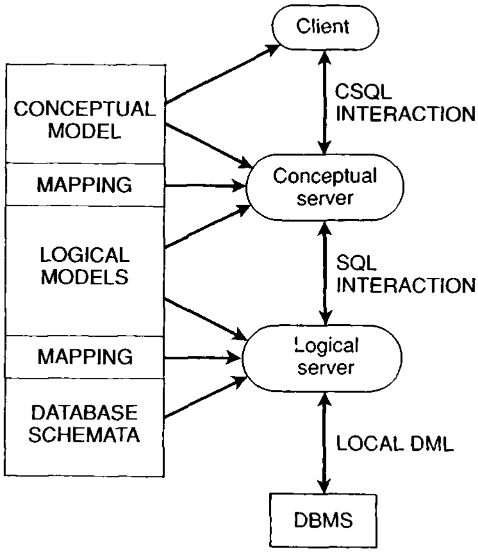


Fig. 3 RIS – use of models

The CIS shares the conceptual model for a partition with application clients and interactions between them are in terms of this model and in the RIBA conceptual data manipulation language known as CSQL (see 3.2).

The CIS shares one of the logical models with each of the LIS within the partition. Interactions between the CIS and a LIS are in terms of the shared model and expressed in a subset of standard SQL.

The CIS employs the mapping between the conceptual and logical models to first translate clients messages from CSQL to SQL and then to select one or more of the logical servers as necessary.

The logical servers have access to a logical model and the schema for the database to which they are connected. The mapping between them is used to translate messages from the CIS, in SQL, to the language understood by the database management system. For relational databases this translation is relatively trivial.

3.1.1 Conceptual Model The basic model, used by the CIS, has the form of a directed acyclic graph whose nodes are sets and whose arcs are binary relations between the sets. A number of interpretations of this graph model are possible. In the present implementation of the RIBA information service it appears externally as an EAR model though there are plans for an alternative, object oriented, interpretation in the future.

Thus:

- the nodes of the graph are named entity, relationship and attribute sets.
- attribute sets are associated with presentation types such as 'string', 'date', 'integer' etc.
- arcs of the graph are labelled as 'has-attribute', 'isa', 'identified-by' etc.
- the arcs are directed from entity towards attribute, relationship towards entity and towards specialisation.

Figure 4 is a simple example taken from the RICHE project.

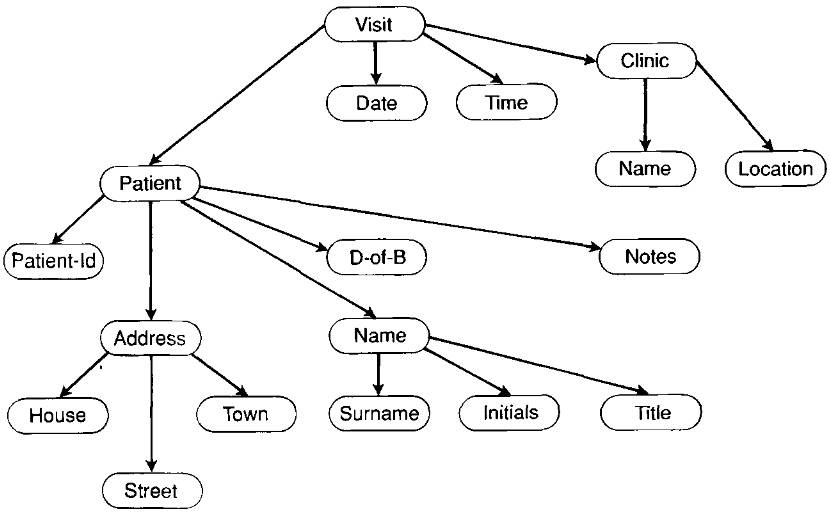


Fig. 4 RIS — information model example

3.1.2 Logical Model The logical model is record based.

The underlying databases are assumed to hold named sets of records (record types) which in RIBA are known as *fragments*.

The description of a fragment assumes that a record may be divided into subrecords that are either ordered or self-identifying and that each of the subrecords contains fixed length and, at most, one variable length field. The fields are named and one or more of the fields may be designated as forming a prime key or record identifier.

This record model maps easily onto a wide range of data structures from relational tables through Codasyl records to document formats and from the point of view of the conceptual server is a simplifying unification of the potential variety in the database schemata.

3.1.3 Mappings Each of the items of a conceptual model is mapped to one or more of the logical fragments. Attributes are mapped directly to fields within fragments. Entities and Relationships are not mapped to fields but are explicitly associated with fragments to which, for example, their identifying attributes are mapped.

In the simple case each of the fragments maps to a named table or record type in a database schema.

More generally the content of a fragment may be separated over more than one database, on some geographical or temporal criterion. In this case the predicate for record instance inclusion is supplied as part of the mapping. For example — a fragment called Account may be mapped to the Accounts table in all of the branch databases of a bank using the predicate "if SortCode='nnnnnnnn'".

3.1.4 Data Description Language (DDL) The following examples are intended to give a feel for the data description language that is used to generate all the models and mappings. In practice the description in DDL is a product of a modelling tool used by information analysts and provided as part of the RIBA toolkit.

The description for each partition contains first of all statements that define the graph model in its EAR interpretation.

E.g.
Person has date Date-of-Birth, virtual Address, virtual Name, optional Telephone, Status;
Person identified by unique National-Insurance-No;
Address comprises House, Street, Town, PostCode;
Name comprises Title, Surname, Initials;
....

These statements are followed by those that give access permissions for named user-groups.

```
E.g.
group Doctor read all;
             write status;
group Nurse read all except Status;
            write none;
```

At the logical level statements name the fragments and declare their content.

```
E.g.
fragment Persons;
Person;
Name;
Address;
NINO=National-Insurance-No 0 12 P;
TI=Title 0 6;
SN=Surname 0 30;
IN=Initials 0 6;
....
end of fragment;
```

By way of explanation – National-Insurance-No is mapped to a field called NINO which is contained in the first and only subrecord, has a width of 12 bytes and is the primary key field.

The fragment definitions are followed by the mapping of fragments to databases.

```
E.g.
Persons contained in Live-Records as P-Table if Status='Live';
Persons contained in Dead-Records as P-Table if Status='Dead';
```

Finally there are assignments of databases to Database Management Systems.

```
E.g.
Live-Records in Patients at Node-1, Oracle, Lock=Row;
Dead-Records in Archive at Centre, Ingres, Lock=Table;
```

3.2 Data Manipulation

The application clients converse with the CIS using a language known as Conceptual Structured Query Language (CSQL). This is basically a small subset of standard SQL from which all logical references are removed.

That is:

- no From or Into clauses,
- no Table-name prefixes,
- no explicit joining of tables in Where clauses.

Some examples of insertions, queries and updates follow.

Insertion

"insert (Customer-No, Account-No, Account-Name, Overdraft-Limit) values (895362,90291403,L.E. Crockford, 0)"

This would be converted by CIS into a set of SQL statements –

"insert into ACCOUNTS (ACC-NO, ACC-NAME, OVRLIMT)
values (90291403, 'L.E. Crockford', 0):
insert into CUST-ACCOUNTS (CUST-NO, ACC-NO)
values (895362, 90291403)"

– that would be dispatched to logical servers as an atomic transaction.

Queries

Queries are performed using a virtual cursor.

"declare C1 select Surname, Balance where Balance>1000000
open C1
fetch C1
....
close C1"

In this example the "declare" causes CIS to use inferential techniques to determine a suitable path to navigate through the underlying databases and to form a chain of tasks expressed in SQL with explicit linkages.

The "open" forms a virtual table with columns for Surname, from a customer detail database perhaps, and the Balance from an accounts database.

Each "fetch" causes the SQL chain to be executed and re-executed in order to complete and return a row of the virtual table, or to respond "out of data" if the first task in the chain returns nothing.

In some circumstances, where there is ambiguity, additional phrases are employed – e.g. "via Customer-Mortgage" (for resolution of multiple relationship choice), "Date of Delivery" (where there is overloading of names like Date).

Update

"declare C1 select Surname, Initials, Appointment-time where Clinic= 'ENT' for update"
open C1
fetch C1
update C1 set Appointment-time='10:45'
....

Update requests are performed within a query cursor that has been declared "for update" and they are issued after a successful "fetch". That is, the update affects the current row of the virtual table.

Since the virtual table may be derived from many database tables a single CSQL update request may, like "insert", be translated into many SQL updates to be performed concurrently as an atomic transaction.

Delete

Deletes are performed, like updates, following a successful fetch within a cursor.

It is not assumed that it is the intention to delete all the underlying records from which the virtual record was formed. The CSQL "delete" must explicitly identify what is to be deleted by giving the names of entities or their identifying attributes.

For example if a "fetch C1" from a virtual table containing attributes of Customer, Account and Customer-account is followed by "delete C1 Account" then the underlying account record will be deleted together with the customer-account relationship. If "delete C1 Customer, Account" then the underlying customer record will also be deleted.

The maintenance of referential integrity is currently the responsibility of the application though an enhancement is planned that will deny the deletion, in this example, of the customer-account relationship only.

3.3 RIS Implications

Experience with the current implementation of RIS suggests that several important benefits may accrue for its users.

> The program developer's view of information is very close to the business information model and is not complicated by concepts that are particular to any database technology, e.g. relational tables or owner coupled sets.
>
> The development of an application can proceed as soon as the conceptual model of its information is available and is independent from database design and administration.
>
> Information can be relocated or migrated between different database management technologies without impact on the application code.
>
> There is a resultant reduction in program code size and complexity.

The generality of the approach that has been adopted should allow support for new modelling concepts and techniques.

4 RIBA Future

ICL is continuing the development of the platform facilities and the various transparency mechanisms are being enhanced. New developments are addressing system management issues, security, high performance transaction processing, distributed application generation tools and so on. The range of operating systems on which the platform runs is also being extended and will include the ICL VME operating system.

Several collaborative projects have been set up to implement pilot operational systems which will validate the RIBA technology against real business requirements.

One of these, the OASIS project, brings together Scottish Hydro-Electric PLC, the IT Institute of Nottingham University, GID Ltd and ICL. The partners are enhancing and applying the RIBA technology to meet particular IT needs of the electricity industry. The project is developing a distributed Customer Information Service within which the new technologies of Open Distributed Processing will coexist with the older systems on which the electricity industry depends in its business. OASIS will not only demonstrate the applicability of ODP to commercial data processing, it also aims to prove an evolutionary approach to systems modernisation and development. OASIS is being conducted as one of the projects supported by DTI under its Open Distributed Systems Architecture (ODSA) programme.

Acknowledgements

The authors would like to acknowledge the excellent work and assistance provided by a number of projects and organisations:

The ANSA project, Cambridge
The ISAD project in Bell Northern Research Laboratories, Harlow
The Department of Trade and Industry for financial support under the ODSA programme
The Commission of the European Community for financial support under the ESPRIT programme

References

[ANSA89] *ANSA Reference Manual* (01·00), APM, Cambridge, March 1989
[ANSA91] *ANSAWare 3.0 Implementation Manual* (RM·097·01) APM, Cambridge, Feb 1992
[RIBA92] *RIBA 2·0 System Guide, OASIS project*, ICL (TEC002), May 1992
[DRAH90] RICHE – A. Drahota, *ICL Technical Journal* (Vol. 7, NO2 pp. 286, 306, 1990)
[ISO89] *ISO Framework of abstractions* (ISO/IEC JTC1/SC2/WG7, 1989)
[MARS91] ANSA – A model for Distributed Computing, D.S. Marshak, *Network Monitor, Volume* 6 *Number* 11, *November* 1991

Biography

Len Crockford

Len Crockford is a senior designer/consultant, from ICL Product Operations, responsible for the specification of the RIBA Information Service. He has over 30 years computing experience, and some of his most significant work has been in connection with the Content Addressable File Store (CAFS). He was a member of the original design team in 1970 and, having helped to build the machine, was responsible for the Associative Data Management System (ADMS) which became the basis for much of the subsequent CAFS software. For the past 6 years, he has acted as a consultant on data modelling and management; participating in three ESPRIT projects – DOEOIS, Bank '92 and RICHE. He is an author of published papers on Design Automation and ADMS.

Tony Drahota

Tony Drahota is a member of the Technical Directorate of the ICL Government and Major Companies division. He has 15 years of computing experience gained in various technical and project management roles within ICL. For the past 3 years he was the ICL technical manager for the Bank '92 and RICHE ESPRIT projects. Previously, he managed the development of Management Support products for use within Government Departments.

Information Technology: Support for Law Enforcement Investigations and Intelligence

S. Blair Southerden
Manager, Investigative Systems, ICL Decision Support Product Centre, Bracknell, UK

Abstract

This paper describes the background to criminal intelligence analysis techniques within law enforcement organisations. Technology now permits the adoption and exploitation of these techniques using ICL's Intelligence Analyst Workbench to provide a knowledge administration system for senior investigating officers.

1 Historical Background

The key issue in the United States during the late 1960's was the threat posed by Organised Crime; the Mafia. A response by the Federal government was an injection of funds into law enforcement, including the commissioning of research to develop techniques for analysing intelligence gathered about criminals and organisations.

The same period in the United Kingdom saw the amalgamation of many of the City and Borough forces into larger County forces, as a consequence of the Royal Commission into the Police which reported in 1962, and the Police Act 1964. Unit Beat Policing, the 'Panda' car and the advent of personal radios were features of policing at the end of that decade. As part of the new structure, police forces created the post of 'collator' to gather information from the patrolling officers, to liaise with CID and to maintain local records on resident criminals.

As part of the organisational reviews which followed the Royal Commission, the Association of Chief Police Officers (ACPO) prescribed a three tier structure for the intelligence process to embrace station, force and regional levels. In 1976 a project known as the Thames Valley 'Collator' experiment was established to identify how computers would help the management of information used by the collator. From then, and through the early 1980s several forces embarked on projects to computerise information about crimes and criminals. Different software tools were used based upon structured or free text retrieval packages, but no single technology dominated – perhaps because some of the requirements went beyond that which technology was capable of delivering.

In 1983 the first UK officers went to USA to be trained in the methods of *Criminal Intelligence Analysis* (CIA) (Peterson, 1990) which had developed from the Federal funding programme of the early 1970's.

To bring the organisational developments in intelligence activities up to the present day (May, 1992), the National Criminal Intelligence Service was formed this year. It incorporates the existing Regional Criminal Intelligence Offices (the third tier) with national units such as the National Drugs Intelligence Unit and the National Football Intelligence Unit. Its staffing includes officers from HM Customs who work particularly on counter narcotics activities. This introduces a further dimension, the multi-agency nature of the task.

A major activity carried out by the police service is investigation, indeed it is the major function of the Criminal Investigation Departments of the service and comprises about 12–15% of all officers. Cases which had a significant impact on British policing were those undertaken in the north of England in the late 1970's and early 1980's into the series of murders labelled by the press as the 'Yorkshire Ripper'. Issues to do with the management of 20 murder investigations (not all of which were finally attributed to Peter Sutcliffe) led to a review by Her Majesty's Inspector of Constabulary for the region. The review concluded that the following lessons had to be learned:

— standardisation of procedures
— computerisation of records
— training of senior police officers
— appointment of an advisory team
— use of specialist and scientific support (Doney, 1990)

The Association of Chief Police Officers established a Working Party on Major Crime Investigation, tasked with developing and implementing the review's proposals. (ACPO, 1983). This working party created the Standard Administrative Procedures, the manual processes adopted by all forces for the management of subsequent investigations. These procedures included the establishment of an 'Intelligence Section'

> "In order that the Senior Investigating Officer can more readily keep abreast of the progress of an investigation, validate and put in priority, potential lines of specific enquiry and assess the requirements for the investigation in terms of manpower and equipment, an intelligence section set aside from the minute by minute running of the investigation should be formed.
>
> ... the staff should be vested in the art of analysing and presenting intelligence."

"Intelligence" (as in criminal intelligence analysis) may be defined as a process of adding value to information through collation, analysis, forecasting and dissemination, with the object of developing the most valid meaning from the information (Harris, 1976). In practice such an Intelligence Section is rarely assembled, due to resourcing constraints and the lack of persons skilled in the techniques. More commonly a Researcher within the incident room staff is assigned the functions of database interrogation, quality assurance and chart preparation.

The HOLMES (Home Office Large Major Enquiry System) specification lays down the requirements for a computer application to assist with the administration of large investigations. (Eagle, 1988) It provides a database of information on people, addresses, vehicles, telephones, events and categories: additionally it provides a module for managing the documents and actions raised in the course of the investigation, and a free text retrieval capability for searching the contents of documents. It does not address the requirements later identified (Oldfield, 1988) for administering the knowledge base of an investigation.

2 Oldfield's Research

Oldfield is a scientific officer at the Home Office and had contributed to (Eagle, 1988). His research was carried out in 1988 to identify the benefits to major crime investigations from the integration of criminal intelligence analysis techniques with existing investigation procedures. A survey of forces indicated that 50% (26 forces) had used CIA during major crime investigations, but that the ACPO Standardised Administrative Procedures lacked, in practice, a formal procedure for handling the knowledge generated by the processing of information during an enquiry. In effect, the problem for the senior investigating officer is "knowing what we know".

The study draws three conclusions as to the reasons why CIA currently failed to meet all the requirements of the police in support of investigations.

2.1 There are particular requirements for a system to manage knowledge acquired in a criminal investigation that are currently not supported by the CIA process.
2.2 The present level of integration of CIA with the ACPO Standardised Administrative Procedures precludes effective use of some of the existing CIA techniques.
2.3 Analysts cannot perform efficiently with the present level of technology available to them.

The report makes three recommendations following from these considerations:

2.4 Solutions must be sought for those requirements of the knowledge administration system that are not addressed by current CIA techniques.
2.5 If solutions can be found then, for greatest effectiveness, the 'improved' processes must be integrated with the standard procedures.
2.6 Finally, work must continue with the development of tools to maximise the efficiency of the analyst.

3 Criminal Intelligence Analysis

As mentioned already, the threat from the MAFIA led the US Federal Government to fund research into law enforcement. Coordinated on Operational Research lines, this resulted in a set of techniques permitting analysis of the unity between criminal and organisations under investigations, which techniques are summarised in what follows.

The successive steps taken within an intelligence department are:

Collection or receipt of information: a large volume of paper based information will be received, either solicited or unsolicited, and this has to be read, evaluated, acted upon or filed. Specific information will be sought on a day-to-day and hour-to-hour basis, using all the media available, computers, telex, fax, telephones, as well as covert sources.

Evaluation: information received will be checked, both for reliability of the source and for accuracy and credibility, by seeking other information which supports or negates the truthfulness of the new information.

Information storage: whatever the storage medium, information has to be registered, indexed, cross-referenced and if, of a time-critical nature, flagged to the appropriate individual for some action. The nature of the information sources emphasises the large amounts of raw text, but increasingly the media includes photographs, video, audio and electronic media. However the information is stored, it has to be made available automatically to analysts for rapid retrieval on demand according to its content or topic.

Analysis: an all-embracing term which covers a number of techniques and styles. The original focus on Organised Crime led to the employment of techniques based upon Social Network analysis. The recent emphasis on narcotics trafficking with the consequent generation of large sums of money has caused law enforcement to adopt techniques for analysing undisclosed incomes and commodity flows. Both intelligence and investigative organisations analyse events across time.

Dissemination: the process is worthless if the product of it, the forecasts or conclusions, are not passed on for action. This may occur through the production of an intelligence report, produced either on an ad hoc basis for a very restricted readership, or as a regular (annual) publication with a wide and public circulation, or as an oral briefing.

However this variety of available techniques has not been integrated so as to be usable in a flexible and coherent manner and their users have tended to be intolerant or suspicious of technical people. Nevertheless the technologies have been applied innovatively and have yielded solutions showing advantage. The opportunity to integrate the tools more closely with the source data will maximise the advantages which can be gained.

4 Crime Investigation

Two useful references about crime investigation are (IACP, 1989) and (Osterburg, 1992) which describe the detailed processes to be followed when investigating specific types of crime, ranging from homicides, other crimes of violence, burglaries, frauds, etc. A significant factor which affects murder investigations in Great Britain is their relative infrequency (704 recorded in 1991, representing 1·28 per 100,000 of the population), compared with, for example, the United States of America where over 24,000 were recorded in the same year (over 10 homicides per 100,000 of the population).

Osterburg (1992) confirms that many studies have shown a prior relationship exists between the victim and killer in about 80% of cases. However there has recently been a significant drop in the clearance rate in USA, i.e. the number of cases in which an arrest is made, from over 80% in 1970 to 66% in 1991. The clearance rate in Great Britain remains consistently around 90%.

The relative infrequency of homicide in Great Britain, coupled with the natural fear of that crime causes the police to treat these investigations exceptionally, and to allocate additional detective resources to those cases. Typically murder investigations may have from 10 to 100 officers attached to the team, at least during the initial weeks when information is most likely to be available and when public attention is focussed on the inquiry by the media. Large numbers of officers generate high volumes of information for the incident room team to process. The role of the senior investigating officer includes the effective, efficient and economic management of his resources and setting the strategy for the investigation by identifying lines of enquiry.

At the preliminary stage of a murder investigation the steps to be taken include the following.

- Recording the crime scene.
- Collecting and preserving physical evidence.
- Identifying the victim.
- Establishing the cause and manner of death.
- Ascertaining a motive for the crime.

- Seeking additional information from those who knew the victim, examination of records kept by the victim, by comparison with police records of similar crimes, and obtaining evidence for forensic examination from possible suspects.

By their nature, murder investigations may be complex; however, other criminal investigations such as fraud, organised crime or terrorism can pose comparable or greater problems. Less serious cases may have their own difficulties because fewer resources are available for deployment and management pressures to close the case can be considerable.

5 Intelligence Analyst Workbench

In 1987 an overseas law enforcement agency was considering the procurement from ICL of a Series 39 mainframe on which to run the company's 'Indepol' software package, and planning to develop a database of information about criminals within their jurisdiction. An overview of the type of information held and its structure is shown in Figure 1.

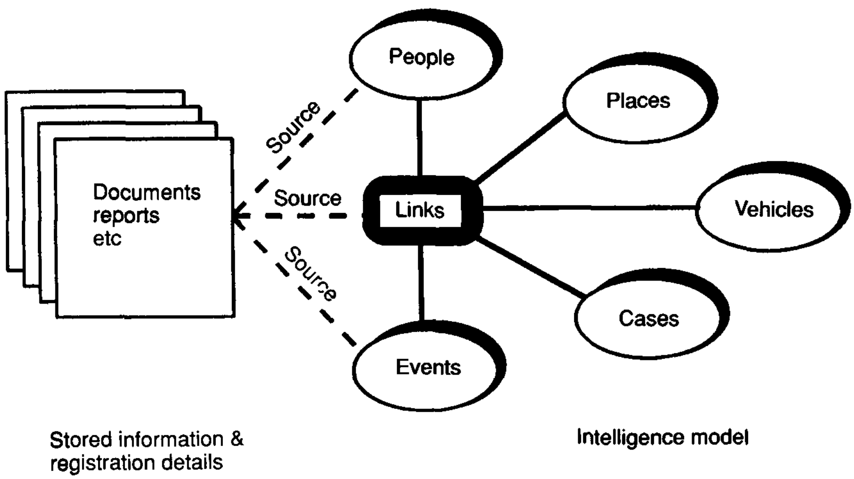


Fig. 1 Intelligence and Investigative Data Models

The criminal intelligence bureau of the force had been reorganised and trained in the techniques of intelligence analysis. As an aid to understanding, the analysts would illustrate their reports with 'link charts', a very simple example of which is illustrated in Figure 2.

The preparation of such diagrams without the support of IT is time consuming, and takes analysts several days. "Why," the customer asked, "given that the data is held on a computer system, can't technology assist in the production of these charts?" This seemed to me not an unreasonable question, given the similarity in the style of display now frequently seen as a component

LINK CHART

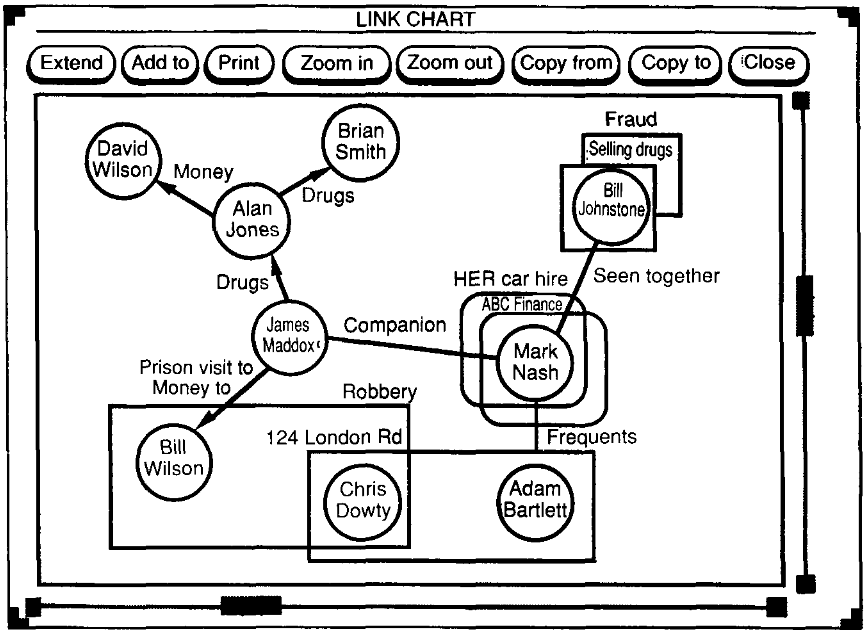


Fig. 2 Link Chart

of CASE tools, or project management packages. But Indepol was not then an environment which was supported by such tools, and of course the functionality that law enforcement was looking for differed from that required by systems designers. Thus the HCI of typical CASE tools was not a model for law enforcement.

A requirements definition was arrived at by ICL through a series of workshops with law-enforcement professionals in the United Kingdom, Hong Kong and Australia. It quickly became clear that obtaining information occurs earlier on in the intelligence process and was a bigger problem for the end user than its analysis. As one Australian put it, "I've got four terminals on my desk at the moment. I haven't got room for another; all the information has to come through a single terminal, irrespective of what computer the data is on." This requirement for access via a single terminal more than any other confirmed the need for a systematic approach to the investigation and intelligence process. Providing a solution to one part of the process alone was unlikely to deliver the required benefits if earlier or more significant problems hindered exploitation.

As a result of the workshops, ICL commissioned the development of a prototype application which demonstrated the feasibility of displaying data relationships graphically, and in a way that end-users could readily exploit.

This research involved a formal collaboration with the Home Office during the period of Oldfield's research. This collaboration has been reported (Oldfield 2, 1988).

6 Platforms

The Intelligence Analysis Workbench (IAW) is conceived as a hardware and software platform which will provide an analyst with an integrated set of software tools for accessing remote databases, extracting sub-sets of data and subsequently carrying out specific types of specialised analysis on that data.

The *hardware platform* was chosen to meet the following requirements.

6.1 Providing a large, high quality graphical display, with the resolution which will permit display of photographic images to a good standard.
6.2 Concurrent processing capability.
6.3 Potential exploitation of AI techniques in future tool development.
6.4 Powerful processor capability for development and end user performance.

It was decided in 1989 that Sun's new SPARC1 workstation best met those requirements. Now, the software runs on a Sun advanced graphics workstation, typically a Sun IPX or a SPARCstation2, with 32 Mb of RAM, a minimum of 400 Mb of disc and a 19″ colour monitor. Printing is provided by a Postscript compatible printer.

The *software platform* is ICL's Decisionpower toolset, specifically the Prolog compiler – Sepia, with Graphicspower providing the graphical user interface (GUI). Decisionpower delivers a development environment which enables rapid prototyping of applications. The nature of the final application was considered sufficiently complex that prototyping was the least risky approach to the development, sure in the knowledge that the end user requirements would change.

Above Decisionpower, the AT&T OPENLOOK interface presents windows, menus, icons and other symbols in place of traditional commands. Multiple windows allow an analyst to display different views of the data simultaneously, allowing comparison and comprehension of information from one or more databases. The contents of these windows can be scrolled vertically and horizontally, and resized, while a "zoom in and out" facility gives additional flexibility.

The product ultimately developed provides a set of tools ready for immediate use by an investigator or analyst using Holmes. For other systems, the software will have to be integrated with the application which holds the investigative data. The scale of the development activity will depend upon the complexity of the application but the effort required is unlikely to exceed two man-months.

7 Applications Software

The software can be divided into two parts: an advanced intelligent terminal emulator and the analysis tools.

7.1 Terminal Emulation

Terminal emulation is required to provide the user access to the host database(s). To date the communications element supports access to the following applications:

Application	Protocol	Connection
Holmes (Unix)	VT220	TCP/IP or RS232
MICA(H) (McDonnell Douglas)	Prism P9	RS232
Indepol (ICL VME)	ICAB02	OSLAN/OSI

Components added to the terminal emulator hold information about the individual application databases allowing:

— descriptive information to be extracted as labels for the icons used in the GUI;
— integrating the keyboard mapping of disparate applications;
— providing a single enquiry interface across multiple applications.

7.2 Analysis Tools

In the environment in which end users will work, new information is continually being added to the database, both in the form of new records as well as updates to existing records. Researching a particular record may involve a series of transactions as illustrated in Figure 3.

This illustrates the retrieval of record A, inspection of that record and subsequent navigation to records B–F, and then subsequent navigation from D to G–I and F to J&K.

Using IAW an analyst opens a ‘new case’, that is to say a new line of investigation or inquiry. The analyst also decides the type of analysis he wishes to carry out. The analyst is likely to want to know all about a particular person, or address, and then to find out what other information is known and which has been linked to that original record. For example for a specific address: who lives there; who visits there; to whom are they

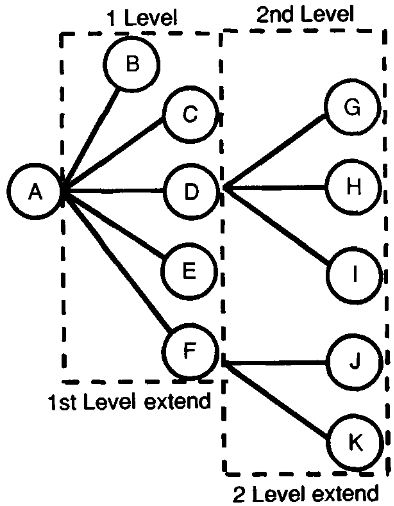


Fig. 3 Extending from a Record

known; and what relationship (if any) do they have with the victim or with a suspect. If one is "researching" the database the most likely choice will be a *link chart format*. The analyst decides on the starting point for the research and enters, for example, the name of the person whose record is to be retrieved. IAW copies the information from the database and displays it in the link chart view. IAW has retrieved for the analyst his 'seed' record A; and then permits him to 'inspect' the record. Inspection displays all of the information held on the record (or as much of it as his security level permits him to see), in a separate window on his terminal.

At this point the analyst can choose to 'extend' from that seed record to retrieve all other records linked to it and, if wished, records linked to that set, and so on. Multiple levels of extension are permitted, though practice suggests that more than five levels probably "over retrieve" for the purpose of analysis. IAW then aids the analyst in laying out the chart by advising him of the types of records retrieved and highlighting those icons to which a record is linked. The term entity is used to refer to a person, address etc. An "entity" may be displayed as a circle, box or 'soft' box. The two box types can contain other entities (circles) to which a physical link exists on the database. The boxes can be 'tiled' for clarity of display. As with a Venn diagram, this technique reduces the number of link lines which tend to clutter a chart. It is particularly helpful in showing people (circles) who are linked to an address or organisation (box), or people (circles) who are referred to in a document (soft box). The development of the software to support the 'box' icons provided interesting challenges as the user requirement specified that when an icon is moved on a chart, all the 'attached' links move with it. A user can therefore move a box containing multiple entities across a chart. Subsequently validating the chart to ensure there are no illegally placed entities is processor intensive, but is a feature much valued by users.

By pointing at individual icons the analyst can read information from each entity and decide whether the record is one which should be *retained*, whether it should be *hidden* (retained but not displayed) or *rejected* from the case. Hieroglyphics are used to prompt the analyst which records have not been inspected or extended from, or have links to entities which have been removed from the chart.

Dummy records and dummy links can be created without compromising the integrity of the main database by storing the information locally on the IAW workstation. Annotation can also be added to the chart for briefing purposes. Hypothesis testing can be carried out, for example by hiding an entity to see the effect of removing an apparently key person from a network. As all the link records associated with that entity will also be hidden, the potential effect on a criminal network can be visualised and assessed.

Further investigations can be carried out in the same way in other windows and subsequently the charts can be merged to provide a total picture, showing, for example, the results of multiple lines of enquiry, or the analysis of sub-groups of criminal organisations.

At any time, hours or days after the development of a chart, an analyst can repeat the 'extend' command from the seed record. A subsequent search of the application database will recover any new links which have been added since the first enquiry to that, or any other entity displayed on the chart. The way the new information is displayed to the analyst ensures that he is aware what it is new. Before IAW the methods that could have been used would have necessitated completely repeating all of the research and comparing the results, a task which becomes longer and more error prone with each iteration.

Two other views which are supported are an Association Matrix and a Network Profile. An Association Matrix (see Figure 4) is an intermediate step used by analysts when developing intelligence without computer assistance. It is a two dimensional matrix which displays all of the entities collated as part of the research with a numerical count of the number of links between any two entities. When network charts are being drawn, the first step is the placing of entities with the largest number of links, so that those links radiate outwards from the centre. IAW displays the entities sorted by number of links and number of entities linked to; there may be multiple links between two single entities. No entities can be deleted from an Association Matrix so it provides a safety check on the research carried out in a particular case.

The final view is a Network Profile an example of which appears at Figure 5.

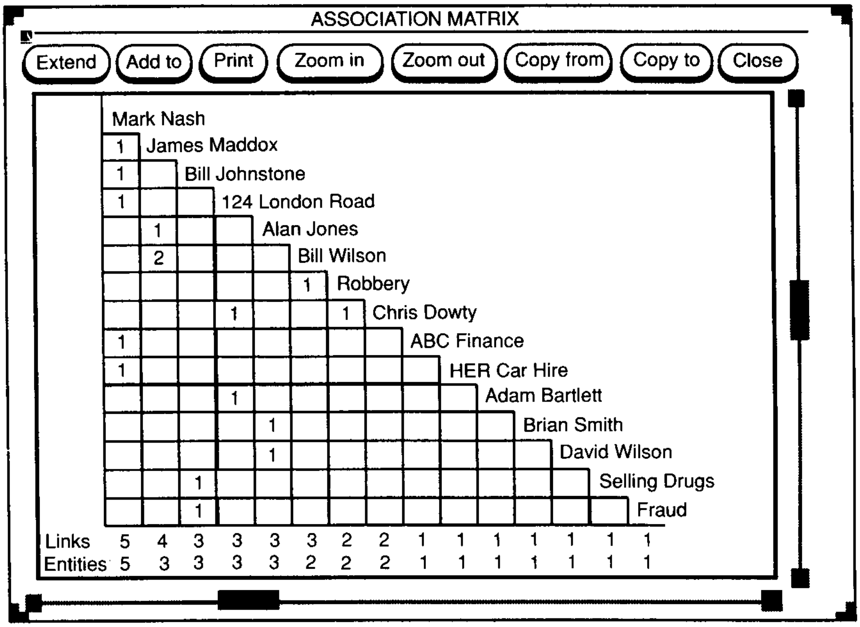


Fig. 4 Association Matrix

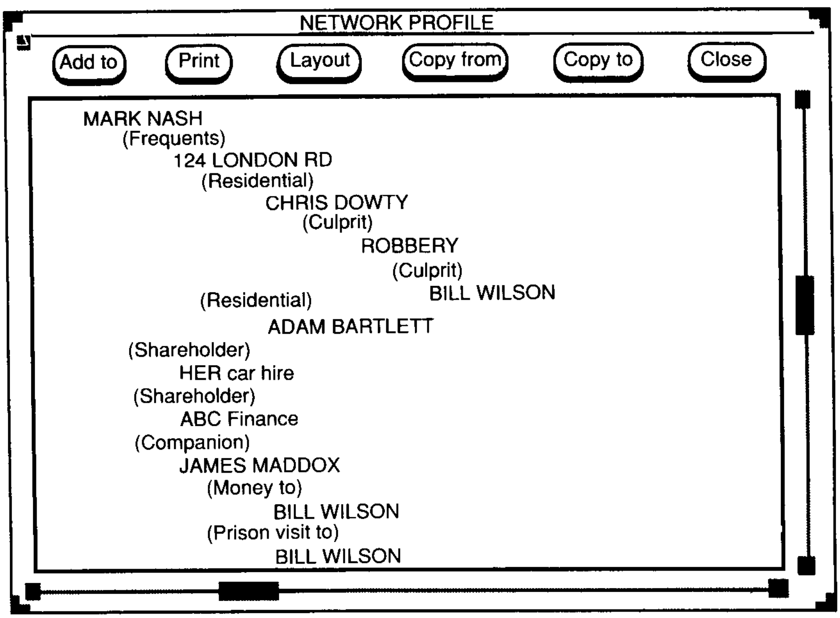


Fig. 5 Network Profile

The Network Profile provides a high level schematic of the information held in an Association Matrix. It is a technique which it had not been thought feasible to develop without computer assistance as it is too time consuming to produce. From the set of records in the 'case' it is possible to select a single record as a start point and to sort the Network Profile from that record, such that linked records are indented according to the number of levels away from the seed record at which they lie. The example in Figure 5 shows that the records directly connected to Mark Nash are 124 London Road, James Maddox, HER Car Hire and ABC Finance. Mark Nash is linked to 124 London Road because he frequents that address. 124 London Road is linked to Chris Dowty as it is also listed as his residence, and so on. In complicated networks where there are perhaps two hundred entities which may have extended six or seven levels from the seed record, it may be vital to investigators to know if two persons are linked directly or indirectly.

8 Conclusions from the use of IAW

IAW has now been used in a small number of investigations where information exists in a Holmes application. The first instances were audits of previous investigations to measure the savings in effort by using IAW as well as seeking to identify any potential new lines of inquiry. In one of the investigations a Victim's Social Network (basically a family tree in this instance) was researched and drawn manually. An experienced analyst took some three weeks to research and to develop the chart. The exercise was repeated with IAW in an elapsed time of three days, and had the added benefit of being capable of easy amendment as the Holmes database is updated. The system has now also been used on an operational investigation with positive benefits to its management.

In particular, Oldfield's conclusions on the use of CIA techniques seem to be supported, as the product of the analysis has provided the senior investigating officer with tools to assist the knowledge administration process. This knowledge is simply not effective when dependent upon techniques unsupported by IAW, as it demands too many resources, and the analysis takes too long to develop. Only through the use of technology can timely analysis be produced.

The IAW platform provides a basis for further collaborative development with law enforcement agencies and software developers. Moreover, the methodology supported by this product may be applicable to other business areas which have databases of related information. Opportunities to exploit these tools are limited only by the imagination of individuals in their exploitation.

References

1 PETERSON, M.P. and RIDGEWAY, R.G. *Analytical Training in Today's Law Enforcement Environment* (Law Enforcement Intelligence Analysis Digest – Winter 1990–91.)
2 DONEY, R.H. *The Aftermath of the Yorkshire Ripper: The Response of the United Kingdom Police Service* published in EGGAR, S.A. Serial Murder: An Elusive Phenomenon (Praeger 1990)
3 ACPO Crime Committee Working Group on Major Crime Investigation, 'Major Investigation Incident Room Standardised Administrative Procedures', *ACPO Crime Committee Publication*, (First report January 1983, Second report April 1985)
4 HARRIS, D.R. *Basic Elements of Intelligence* (rev ed Washington DC,: US GPO 1976)
5 EAGLE, R.W., WILLIAMS, K.C., MYLAM, H.M., WILKINSON, R.A. and OLDFIELD, R.W. Holmes *Application Design Specification* (Home Office London 1988)
6 OLDFIELD, R.W. *The Application of Criminal Intelligence Analysis Techniques to Major Crime Investigation – An Evaluation Study* (Home Office London 1988)
7 OLDFIELD, R.W. *ICL Criminal Intelligence Analysis Workbench Evaluation Study* (Home Office London 1988)
8 IACP., *Criminal Investigation* (International Association of Chiefs of Police, Arlington, Va: 1989)
9 OSTERBURG, J.W., WARD, R.H., *Criminal Investigation: a method for reconstructing the past* (Anderson Publishing Co., Cincinnati, Ohio: 1992)

Further Reading

ANDREWS, P.P. Jr., PETERSON, M.B., *Criminal Intelligence Analysis* (Palmer Enterprises, Loomis, California: 1990)

KIND, S.S., *The Scientific Investigation of Crime* (Forensic Science Services, Harrogate, UK: 1987)

Biography

Stuart Blair Southerden

Blair Southerden joined the Kent County Constabulary in 1967 and served as a police officer for sixteen years. During his service he was seconded to the University of Kent at Canterbury where he read for a degree in Public Administration and Management. These courses introduced him to computing studies and on his return to the force he worked for four years on a project team specifying the requirements for a major crime and criminal information system in conjunction with the Home Office.

He left the police in the rank of inspector and worked for three years with a systems house. He joined ICL as an industry consultant in 1987 and supported the marketing work leading to the establishment of a focussed line of business within ICL Australia. He transferred to his present role to manage the development of the Intelligence Analyst Workbench, a requirement he had identified when working with overseas customers.

Standards for Keyboard Layouts – The Origins and Scope of ISO/IEC 9995

Bruce Paterson
Independent Consultant, London, UK.

Abstract

ISO/IEC 9995, Keyboard Layouts for Text and Office Systems, is a new International Standard due for publication in 1992. It is applicable to the layout of the keyboards of information technology products, including personal computers, workstations, computer terminals, and other devices. It replaces earlier standards for the keyboards of typewriters, adding machines, and telecommunications and data-entry terminals. The paper reviews the main features of the standard, and assesses its achievements and shortcomings as seen by users and by prospective purchasers of keyboards.

As background for the review some highlights of the evolution of keyboard layouts over the past hundred years are given, together with a description of the various distinct user tasks and roles which have influenced the design of keyboards during the era of the electronic computer. Some opportunities for beneficial innovations in keyboard layout are briefly mentioned at the end.

1 Introduction – Historical Background

The information technology products and systems that are available today, when compared with those of thirty years ago, provide a dramatic illustration of the speed at which change has occurred during the intervening time. The remarkable increases in performance, versatility, convenience, and variety have been matched by equally remarkable decreases in the size and cost of comparable products.

Keyboards have been, and still are, the principal means by which user personnel can enter text and data for processing, storage, or communication by IT systems. They have always formed an integral part of IT system installations. But while the general appearance and styling of keyboards, and the internal technology that they utilise, have changed substantially over thirty years the layouts of the keys on successive keyboard products have evolved relatively slowly. In fact the layout conventions used today by the keyboards of IT products have a long history, dating from well before the era of electronic computers.

Before giving a little of that history it is useful first to state what the term "keyboard layout" covers since it appears frequently throughout this paper. A keyboard layout means the physical configuration and relative positions of the keys on a keyboard and the allocation of letters of the alphabet, numbers and other symbols, and associated control functions (such as shift or tabulation) to those keys.

The layout of the alphabetic section of almost all keyboards in use in English-speaking countries today incorporates the QWERTY convention, so-called because this sequence appears along the upper row of letters when reading them from left to right. This convention first came into use over a hundred years ago on one brand of typewriters in the USA, and by the 1890s was already overtaking the alternative layouts that were provided on some other brands. The manner in which each row of alphabetic keys is slightly offset from the row next to it also originated at that time, due to the mechanical constraints of typewriter construction. There is no technical reason to perpetuate it now, and it is retained simply because users are familiar with it. Even the layout of the small rectangular group of numeric keys found on many keyboards today, and arranged as on a pocket calculator, is derived from a key-punch for punched cards that came on the market in about 1900, although at that time the layout was inverted with the zero key in the back row (furthest from the user), keys for 1, 2, and 3 in the next row, and so on.

Evidently the suppliers of typewriters and other keyboard products have recognised over the years that users prefer keyboard layout conventions to remain stable when new models of keyboard become available. Substantial changes of layout have sometimes been introduced when a new model of keyboard was designed to support new types or mixes of task. But in many respects the layouts of the keyboards that are on the market today are the result of a long succession of relatively small changes.

Along with these changes of keyboard layout conventions, over the past forty years or so a succession of related national and international standard specifications for keyboard layouts has been published by BSI and ISO respectively. National standards have also been published in other European countries, in USA and elsewhere. Usually these standards have attempted only to recognise and give a precise description of the principal conventions adopted by current and forthcoming products in the marketplace, but they have enabled a wider harmonisation of layouts between the products of different suppliers to be steadily achieved. The serial numbers of the published standards have also provided a convenient short form of reference to particular layout conventions, as an alternative to the proprietary product references which are so often used for this purpose.

The ergonomics of keyboard operation, and the relationship of the keyboard to the workplace, have since 1982 become the subject of formal standards, beginning in Germany (DIN 66234). An International Standard is now available (ISO/IEC 9241), although the final draft of Part 4 which deals with keyboards was only published in March 1992. For this reason ergonomics aspects are outside the scope of the standard which is the main subject of this paper.

2 User Roles

By the 1960s several different types of keyboard were in use in offices, information processing installations, and elsewhere to support the sharply differentiated tasks that users were required to do. Keyboard operators were usually specialists in the use of one particular type of keyboard, and were trained so that they could achieve high productivity in their assigned task. The principal types of keyboard in use included typewriters (mechanical and electric), data-entry devices (for punched cards and other media), accounting and calculating machines (some for one-hand, others for two-hand operation), and telex and teletype machines for telecommunication of messages. Teletype machines were also in use on computer control consoles, and were increasingly provided for users of multi-access computer systems who were typically scientific and technical specialists or programmers; however in these roles their users generally did not need such a high productivity of keyboard operation.

Naturally the published standards for keyboard layouts made allowances for these distinctions of user role. In consequence, during the 1970s several different standards were published by ISO covering the main types of keyboard layouts that the ISO keyboards committee considered would continue in widespread use during the coming years. The principal distinction made by the standards was between two-hand keyboards and one-hand keyboards. Within two-hand keyboard layouts a further distinction was made between layouts for typewriters and layouts for data-entry using the ISO 7-bit (95-character) code for information interchange. Layouts for one-hand keyboards were intended for adding and calculating machines. The weak point of most of these standards, especially those for two-hand keyboards, was the lack of specification of control key functions suitable for use in information processing tasks, and of corresponding positions for them in the keyboard layout.

2.1 Typewriter Layout

These standards each cover a carefully restricted scope, specifying some significant features of the layout but not the layout as a whole. For example Part 1 of BS 2481 for Typewriters (last reissued in 1982) specifies a layout of between 44 and 48 character keys, in which the positions of the letters of the alphabet, the digits 0 to 9, and eighteen specified symbols and punctuation signs are fixed, but up to sixteen further symbols may be more or less

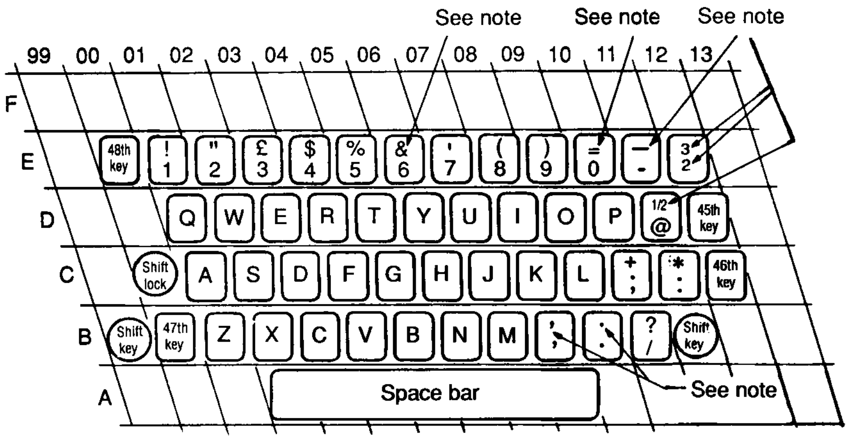


Fig. 1 Character Layout for Typewriters from BS 2481
Note: The indicated characters may be replaced by others when required

freely chosen for the remaining places on the keys (see Figure 1). The positions of shift keys and carriage return are specified in Part 3, but other controls such as backspace or tabulation are not mentioned. Part 2 specifies keytop symbols for twenty different control functions any of which might be present on a particular keyboard, but gives no indication of where the keys for these functions should be placed relative to the layout of the character keys.

Keyboards conforming to this standard will therefore not all have identical layouts, although the most frequently used characters and controls will always appear in the same relative positions on all conforming keyboards. The user must discover the placing of any additional characters and controls on a keyboard by observation whenever necessary. This arrangement was intended to provide a compromise between the need of users to achieve a high productivity of keyboard operation, and to maintain that productivity when they move from one model of keyboard to another, and, on the other hand, the need for some variety in the repertoire of characters and controls on different models of keyboard intended for different price brackets in the market or different categories of document content.

2.2 *Data Entry Layout*

No standard was published for the widely used data-entry keyboard layout, based on the IBM Model 029 card-punch convention. This layout provided one-hand and two-hand operation on the same keys and a limited character repertoire with capital letters only. Because it was intended for tasks in which a high productivity was demanded, suppliers generally avoided providing variant layouts which might impair productivity, and a standard was considered to be unnecessary.

2.3 National Layouts

Also during the 1970s, and subsequently, national standards were published in many countries to specify a precise layout for all 95 characters of their national version of the ISO 7-bit code. These layouts typically included positions for the commonly-used accented letters in the national language, and sometimes also for separate accents intended to be combined with letters when required. In countries where the national language uses a different alphabet or script, such as Arabic, Cyrillic, or Greek, national or regional standards for corresponding keyboard layouts were also published.

3 Visual-Display/Keyboard Devices

The availability of these ISO and national standards helped to bring about a considerable measure of harmonisation of keyboard layouts on the new VDU/keyboard computer terminal products that were then coming on the market from many different suppliers. By this time such keyboards often included a two-hand alphanumeric block of keys and a separate one-hand numeric block to the right of it, mounted on a single base. For each of these blocks of keys suitable standards were available for adoption with little or no deviation. But initially there was little experience of the many and varied ways in which VDU terminals, interacting with computer application programs, might be used. Therefore the control functions provided, and their positions within the overall keyboard layout, were correspondingly diverse among the products that appeared on the market. The choice depended on the judgement of the individual supplier as to the user roles and tasks that the terminal should support and the layout conventions which those users might already be familiar with.

3.1 Variety of Control Key Layouts

As the potential of VDU terminal devices became more broadly recognised and exploited from the mid-1970s onwards, a bewildering variety of keyboard layouts could be seen among the numerous new products that emerged. Some products were designed as dedicated word-processors; their layouts were derived from the typewriter standards and conventions. Other products were intended as general purpose computer-access terminals; their layouts were derived from the adding-machine and the ISO 7-bit code layout standards. But many models of keyboard incorporated large numbers of additional control keys, arranged in blocks at either side of the alphabetic or numeric blocks, or in a long row across the top. These keys were allocated to a great variety of special functions, some associated with direct editing of data on the VDU screen, and others for interacting with the supporting applications in local or remote information processing systems. In these circumstances the user, once familiar with a particular keyboard model, might have to spend a significant amount of "learning time" before attaining a similar level of proficiency on a keyboard from a different supplier, on which the layout of control keys might be different.

The great variety of keyboard layouts tended to limit the choice of product for use in high-productivity tasks, and some purchasers chose to buy all their keyboards from suppliers who could offer layouts identical to those of the keyboards already installed, to ensure uniformity throughout a department or organisational unit. However increasing numbers of users by now were using VDU terminals for applications where speed of keying was not such a critical factor. These users could allow themselves a little time to look at the positions of keys during normal use of the keyboard, but would still benefit from the fact that different models of keyboard adopted similar layouts for the character keys, and also for the block of numeric keys when provided. Thus the lack of standards for the control key positions was for some users a relatively minor inconvenience.

3.2 Personal Computers

The number of keyboard users, and the variety of applications that they made use of, grew rapidly during the 1980s following the arrival of the "personal computer" and its widespread takeup in offices of all kinds. By this time it had become clear that the majority of control functions needed by a keyboard user are quite specific to the individual application that he is using at the time. When in addition a keyboard can be used for many different applications in turn, perhaps from different suppliers of application software, the allocation of such functions to particular keys clearly cannot all be known in advance nor marked on the keyboard during its manufacture. Instead the handbook for the application became the place where the allocation of control functions to key positions was described. A keyboard overlay to label the allocations might also be provided. Many suppliers of keyboard products concluded that in this situation convergence towards a common convention for the overall arrangement of keys on keyboards would be beneficial for all concerned, even though not all keyboards would be able to carry application-specific markings. The opportunity now existed for agreement on a new ISO standard to cover the layout of the modern office-system keyboard as a whole, and at the same time to bring up to date the existing ISO standards which were no longer properly applicable.

4 Outline of ISO/IEC 9995

In April 1986 the ISO keyboards committee, which had been relatively inactive for several years, was instructed to review the existing ISO standards for keyboards to determine if they were still valid, to assess if harmonisation was achievable among them, including removal of the distinctions between text applications and data applications, and to advise on the requirements for new work items in this area. The work was soon under way and the outcome was a new and comprehensive standard for keyboard layouts, ISO/IEC 9995 Keyboard Layouts for Text and Office Systems, which was approved in June 1992 and is due for publication later in the year.

With the publication of ISO/IEC 9995 a standard directly applicable to the keyboard layouts of today's information technology products will at last be available. It aims to cover the field as widely as possible, including personal computers, workstations, computer terminals and so on. It aims to provide a single harmonised standard for the keyboards of IT products and of typewriters, replacing the previous standards for typewriters and adding machines. It also provides harmonisation with the numeric keypads of other products in everyday use such as calculators, telephones, and automatic bank teller machines.

The new standard will allow some flexibility in the overall configuration of a keyboard layout, and while fixing some details of layout quite strictly will allow a certain degree of freedom in other details. As with the previous ISO standards which are to be replaced, this standard aims to provide a compromise between the varied needs of different users. It provides commonality of layout for frequently-used keys, so that users' skills will still apply when they transfer between different keyboards, but at the same time it allows freedom for different models of keyboard to be produced to suit different classes of application, for example multi-function workstations or compact dedicated word-processors.

4.1 Main Sections of the Layout

The standard identifies four main sections within a keyboard layout, any or all of which may be present. The arrangement is firmly based on the keyboard layout of the popular personal computers now on the market or already installed (see Figure 2). The sections defined are a typewriter-like alphanumeric section, a calculator-like numeric section placed to the right of it, a small rectangular editing section placed between them, and a function section comprising one or more rows of keys across the back or at the lefthand side (or both) of the rest of the layout. Each of the sections is regarded as consisting of a central core of keys, known as zone zero, surrounded by other zones which may include keys for supporting or related functions (see Figure 3). In general the layout in zone zero of a section is specified more tightly than in the surrounding zones.

Each of the four main sections of a keyboard layout is specified in a separate part of the standard. There are additional parts to specify the overall configuration and general principles, the symbols that may appear on keytops, and the layout of a character set to supplement the Latin alphabet for multilingual use. The parts are numbered as follows.

1 General Principles Governing Keyboard Layouts
2 Alphanumeric Section

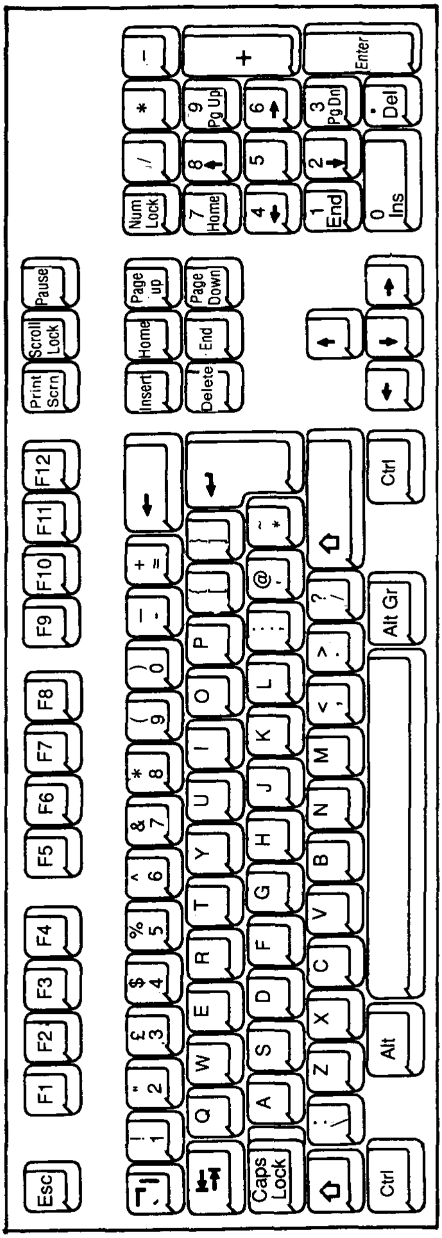


Fig. 2 Typical Keyboard Layout Conforming to ISO/IEC 9995

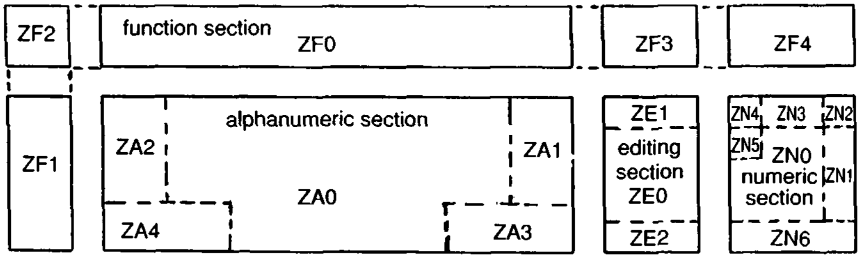


Fig. 3 Arrangement of Keyboard Sections and Zones from ISO/IEC 9995

3 Common Secondary Layout of the Alphanumeric Zone of the Alphanumeric Section
4 Numeric Section
5 Editing Section
6 Function Section
7 Symbols Used on Keyboards to Represent Functions

A conforming keyboard must meet the requirements of Part 1, and of all other parts which are relevant to that particular model of keyboard. Claims of conformance must identify which parts of the standard are met.

4.2 Features Not Specified

The allocation of the principal set of characters to the keys in the alphanumeric section is not specified in Part 2 of the standard, but is assumed to be specified by a national standard or convention. Any claim of conformance that lacks a reference to such a standard will leave one of the main features of the layout undefined. Within the editing section the only functions that are identified and allocated to keys are the four cursor controls, although some additional recommendations are made. Within the function section only the Escape function is allocated. The allocation of all other keys in both of these sections is assumed to be defined by the application that is in use at the time, and is not specified by the standard. The functions of these keys may be identified by means of corresponding legends or symbols on the keytops. Alternatively non-specific legends may be shown, such as the legends F1 to F12 commonly seen on personal computer keyboards. The specific functions for each application must then be identified in some other way.

It will be clear from the preceding outline that ISO/IEC 9995, in conjunction with related national standards, goes a long way towards promoting a desirable degree of harmonisation between different models of keyboard. Users will increasingly benefit from its provisions as non-conforming models of keyboard already installed become obsolescent and are taken out of use. However the standard on its own will generally not meet the needs of the staff in a purchasing department when they wish to issue a purchase specification for keyboards that are identical, or very similar, to those already in use in the organisation. Reference to the relevant parts of the standard will not provide a sufficiently precise specification for that purpose.

The lack of precision achieved within some parts of the standard is distinctly disappointing. The standard would be of more immediate benefit to users if it included a small set of fully-defined layout options. In fact most parts suffer from an excess of options of detail, with no convenient method of referring to them. It is the outcome of a drafting process in which initial hopes of greater precision were not fulfilled. During this process various interest groups tried to ensure that recent styles of keyboard layout would not be treated as non-conforming, and that scope remained for variety and future innovation.

5 The Parts of ISO/IEC 9995

This section gives a brief summary of the scope of each individual part of ISO/IEC 9995, together with an indication of the impact on keyboard users.

5.1 Overall Layout and Principles

Part 1 (General Principles) defines the four main sections of a keyboard layout, as already described above. It defines the various ways of indicating to the user the allocation of characters and functions to the keys, and defines how the characters and other visible symbols should be placed when they are marked on the keys. It requires that when an allocated character or function is not marked on the key the corresponding information should be made available to the user in some other way, such as in an accompanying product description document. It also defines a numbering system for key positions which is used for reference purposes throughout all parts of the standard, although this system is not usually of direct interest to the keyboard user.

A significant innovation in Part 1 is the concept of group shift, to supplement the traditional concept of up and down shifts on typewriters. It enables additional sets of characters (or controls) to be allocated to the keys, and accessed by using a special group-shift key. The group shift can be thought of as a right-and-left shift; when the keys are marked with characters from group 1 (the primary character layout) and from group 2, the group 2 character will be shown to the right of the group 1 character on each keytop (see Figure 4). Within a group there are three levels of shift, instead of the traditional two. Thus up to three characters per group may be allocated to each key. Extra shift keys must be provided to enable these features to be used, e.g. the Alt Gr (alternate graphic) key on PC keyboards. Group shift will be particularly useful for keyboards that support multi-alphabet operation.

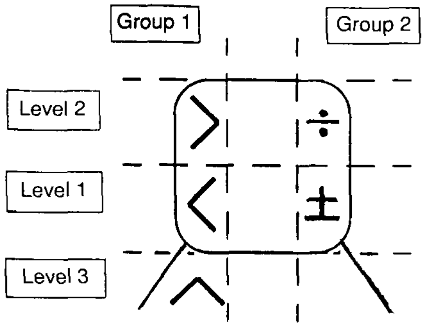


Fig. 4 Example of Keytop Showing Group and Level Shifts

All ergonomic and dimensional aspects of the keys and the layout of keyboards are outside the scope of the standard. They are covered instead by ISO/IEC 9241–4 Ergonomics – Requirements for Office Work with Video Display Terminals, Part 4: Keyboard Requirements. That standard specifies the dimensions of the keytop and the keytop shape, the distances between keys, the distance of key depression and the key force required, the overall slope of the keyboard surface, and other factors.

One aspect of ergonomics deserves a brief mention. A left-handed user might find some tasks inconvenient because the numeric section is placed at the right-hand side of the keyboard. However a separate unit comprising just a numeric block could be provided for such users if necessary, conforming to Parts 1 and 4 of ISO/IEC 9995.

5.2 Alphanumeric Section

Part 2 (Alphanumeric Section) defines the permitted configurations of keys in the alphanumeric section of a keyboard layout. Most keyboards today provide 47 or 48 character keys to accommodate all 95 characters in one of the national versions of the ISO 7-bit code. However Part 2 allows a minimum of 45 such keys, and also allows for possible expansion beyond 48 keys. As already mentioned, the allocation of characters is assumed to be specified by an appropriate national standard, but typical allocations for the letters A to Z and the digits 0 to 9 are given, for guidance when a national standard needs to be newly developed.

Two British Standards for character key allocations are available at present, BS 2481 (1982) for Typewriters, mentioned previously, and BS 4822 (1980) for the UK version of the ISO 7-bit code (i.e. the version that includes the pound sterling sign). Of these standards BS 4822, equivalent to ISO 2530 (1975), was widely adopted in the past for VDU terminals. However in the last few years the keyboards of many models of personal computer sold in Britain have adopted a layout in which many of the less frequently used symbols have been moved to other keys, mostly quite close to their previous positions. This layout is derived from the USA standard ANSI X4.23 (1982). That standard was published after an extensive survey of the opinions of users in many different market sectors. It has the advantage of a harmonised layout based on the long-standing typewriter conventions but also includes all of the characters of the ISO code. The adoption in Britain of the ANSI-based layout virtually eliminates the previous small but irritating differences between the character layouts of American and British versions of the same model of keyboard. It may thus be beneficial in the longer term, although for the next few years coexistence with the BS 4822 and other layouts will be equally irritating to many users. Suitable amendments to BS 4822 are now planned.

Part 2 also defines the positions of the principal control functions such as the shifts, return (i.e. new-line, enter data, etc.), tabulation, and backward erase. Guidance is given for other common functions. The traditional offset between one row of keys and the next in the alphanumeric section is recognised, but the layout does not have to adopt it, nor is the dimensional pitch of the offsets defined. A keyboard in which the character keys were arranged according to a strictly rectangular grid, like the arrangement of the numeric block, would still conform to the standard. Although it would look and feel distinctly odd to most users initially, it might have some ergonomic benefit.

5.3 *Latin Alphabet for Multi-Lingual Use*

Part 3 (Common Secondary Layout) is derived from ISO 8884 (1989) Keyboards for Multiple Latin Alphabet Languages: Layout and Operation. This part defines a layout of supplementary characters for allocation to group 2 shift (described above) on the keys of the alphanumeric section. In conjunction with any suitable national standard character layout for group 1 it enables any character from the 333-character repertoire of ISO 6937 to be keyed in by the user. This repertoire supports forty different languages that use the Latin alphabet. 155 characters in the repertoire are letters of the alphabet with accents or other diacritical marks, and may be entered by keying the accent followed by the letter. While this layout fills a specialist need it is unlikely to prove very popular. Some keyboards on the market have adopted an alternative layout which provides a similar capability but, by reducing the amount of level shifting needed in group 2, on average requires fewer keystrokes.

5.4 *Numeric Section*

Part 4 (Numeric Section) defines the small rectangular block of keys that includes the digits 0 to 9. It defines various alternatives for the symbols and functions that may be allocated to the surrounding keys, to suit different applications such as numeric data entry, general purpose calculator, or telephone number entry. Unfortunately there are two widely-used but conflicting conventions for the arrangement of the digits. In the "7–8–9" layout, adopted by pocket calculators and derived from the ISO adding-machine standard, the keys for 7, 8 and 9 are in the top row (i.e. furthest from the user), but in the "1–2–3" layout specified for the digit keys on telephones (in CCITT E.161) the keys for 1, 2, and 3 are in the top row. Although attempts were made to reconcile this difference, they were not successful. Accordingly Part 4 permits either layout to be adopted on a keyboard, depending on the predominant type of use intended. A keyboard which allows the user to select which layout he wants by means of an associated shift-lock key would also be in conformity. Although Part 4 expresses a preference for the telephone keys convention it is unlikely to cause the calculator convention to fall out of use. Users who rely on one of these conventions in high productivity tasks would need time to adjust if required to change over to the other convention.

5.5 *Editing Section*

Part 5 (Editing Section) defines a small rectangular area, between the alphanumeric and numeric sections, for the keys used for manipulating text and data on the VDU screen. Two alternative configurations are specified for the four cursor control keys, and recommendations are given on the placement of functions Delete, Insert, Next Page, and Previous Page. There is thus great freedom in the choice of functions provided within the editing section of a keyboard, and in their allocation to specific key positions.

5.6 *Function Section*

Part 6 (Function Section) defines a long narrow area above (i.e. beyond) the top row of the other three main sections of the keyboard. It may alternatively, or additionally, extend along the lefthand edge of the alphanumeric section. Keys in the function section are intended for interactions with the application(s) in use, and accordingly Part 6 does not specify particular functions for them. Individual keys may be marked either with application-specific legends or symbols, or with legends that merely act as references to information provided elsewhere, as already mentioned above. The only exception is the Escape function which, if present, must be allocated to the key at the top left corner of the keyboard.

5.7 *Symbols for Functions*

Part 7 (Symbols Used to Represent Functions) defines a collection of over fifty symbols for individual functions which may be allocated to the keys on a keyboard. The corresponding symbol may be marked on the keytop as an alternative to marking with a legend to identify the function. Marking of keytops with symbols has the advantage that the functions can be recognised by users who are not familiar with the language of the user population for which that particular national version of keyboard product was intended. Since in most installations this is not an important consideration, legends for the less frequently used function keys are probably more convenient than symbols, particularly for the occasional user.

6 Discussion

The achievements and shortcomings of the new standard are summarised here, and its utility to users and prospective purchasers of keyboards is briefly reviewed. Finally some opportunities for further standards development are suggested.

6.1 Summary of Features

The disappointing lack of precision of ISO/IEC 9995 has already been mentioned (4.2 above), and will have become more evident from the descriptions in section 5. In summary the aspects of layout that are well-defined by the standard are:

— the four main sections of the keyboard and their overall configurations;
— group shift and level shift, and corresponding keytop markings;
— unique symbols for control functions, and a clear visual distinction between control keys and character keys;
— only two alternatives for the numerals in the numeric section, and for cursor controls;
— a small core set of controls for the alphanumeric and numeric sections;
— the supplementary Latin characters for multi-lingual use.

However the standard does not define:

— layouts for the character keys in the alphanumeric section;
— selections of function keys, and their layout, for common classes of application such as word-processing;
— a scheme of nomenclature by which fully-defined layouts can be readily identified or classified for reference purposes.

6.2 Utility for Purchasers and Users

Claims of conformance to particular parts of the standard will generally not tell a prospective purchaser of a keyboard all that he needs to know about its suitability for his needs. He will have to consider in addition such points as:

— keyboard dimensions, relative to available space on the work surface;
— alphanumeric character repertoire and layout;
— provision and layout of control and function keys, additional to the core sets in each keyboard section;
— need for compatibility, in general or in detail, with keyboards already in use among his user group.

In practice a proprietary product reference will still be the easiest way to identify a fully-defined layout.

Nevertheless conforming keyboards will have many features in common. They are suited to a wide range of applications, user roles, and differing levels of operating skill. When users transfer between different models of keyboard, they will find much that is already familiar. Coexistence of different models in the same installation will not seem quite so daunting. On a worldwide perspective the standard enables conforming keyboards to adopt character repertoires and languages suited to users in other countries.

6.3 *Future Opportunities*

Publication of ISO/IEC 9995 does not imply that the task of standardisation in its field has been completed. Extensions can be envisaged both to specify the relatively undefined features in greater detail and to recognize alternative features or configurations of layout that might offer benefits to users. But the keyboard is only one component in the overall man-machine interface to information technology equipment. At present many other features of that interface are not covered by standards at all. Standardisation of further details of keyboard layouts at this late stage would probably not bring commensurate benefits, and could divert attention from more fruitful initiatives.

Many features of keyboard layouts have evolved slowly over a long period of time. Users can adapt quite quickly to small changes, but large changes require an investment in retraining or familiarisation which must be commensurate with the expected benefits. The new keyboard standard records the main features of today's keyboards; it does not point the way ahead. Ergonomics experts have in the past devised new physical configurations of keyboard intended to reduce the physical strain of keyboard operation, particularly on intensive users. They have also devised rational character layouts to replace the deliberately clumsy QWERTY, with the aim of improving the productivity of trained users or of easing the learning and memorising process for the untrained or occasional user. Because of the difficulty of introducing these new features in an evolutionary way they have not been widely accepted in the market. Today's keyboard technology provides a better opportunity for them. For example the multi-alphabet capability of a keyboard with group 2 shift could be harnessed to provide an optional substitute for the QWERTY layout that the user could switch on or off at will; not exactly evolution, but cautious coexistence. If users welcome such innovations the standards will soon follow.

Standards such as ISO/IEC 9995 represent a consensus of good practice in the IT industry. They are not backed up by any legal sanction. Suppliers will offer conforming keyboards as long as the demand exists in the marketplace. Users who regard the standard as beneficial must remember, or remind their purchasing agents, to ask about conformance to standards when considering new purchases.

7 Acknowledgements

Maurice Blackburn of ICL, Bracknell, was the UK representative on the ISO keyboard layouts committee in the early years of the drafting of ISO/IEC 9995. After his retirement in 1988 the task was taken over by Jim Jackson of IBM UK Limited, Greenock, Scotland. Their extensive contributions to the work of the committee have been highly beneficial to the content and quality of the final standard. The author wishes to thank them for their cooperation in promoting his contributions during the drafting of the standard, ånd for reviewing this paper at the final draft stage.

A slightly modified version of this paper forms a part of the book "User Needs in IT Standards" to be published during 1992 by Butterworth-Heinemann, Oxford. Thanks are due to the publishers for permission to reproduce it here.

Figure 1 is reproduced, in amended form, from BS 2481: Part 1: 1982 with the permission of BSI. Copies of the complete standard can be obtained by post from BSI Sales, Linford Wood, Milton Keynes, MK14 6LE, UK.

References

ANSI X4.23–American National Standard for office machines and supplies – alphanumeric machines – keyboard arrangements 1982.
BS 2481: Part 1: Typewriters, Specification for keyboard arrangements 1982.
BS 4822: Keyboard arrangements of the graphic characters of the United Kingdom 7-bit data code, for data processing 1980.
CCITT E.161 Arrangement of figures, letters, and symbols on rotary dials and push-button telephone sets
DIN 66234: VDU Workstations – Part 6: Design of the workstation (published in German by DIN, Berlin)
ISO 2530: Keyboard for international information processing interchange using the ISO 7-bit coded character set – Alphanumeric area 1975.
ISO 6937: (2nd edition) Information technology – Coded graphic character set for text communication – Latin alphabet 1992.
ISO 8884: Information processing – Text and office systems – Keyboards for multiple Latin-alphabet languages – Layout and operation 1989.
ISO/IEC 9241–4 Ergonomic requirements for office work with visual display terminals – Part 4: Keyboard requirements (draft, March 1992).
ISO/IEC 9995 Information technology – Keyboard layouts for text and office systems (due for publication in 1992).

ESS – a Solid State Disc System for ICL Series 39 Mainframes

A. W. Jenkins
Corporate Systems Division, ICL Manchester, UK

Abstract

Solid State Disc Systems are becoming an essential element in the storage hierarchy of large computer systems. They remove the "access gap" created by the fact that growth in performance of processors has greatly exceeded that of magnetic disc systems. This paper describes the Extended System Storage (ESS), a solid state "disc" system developed for ICL Series 39 systems. The key requirements for the system are analysed and the design features included to satisfy these requirements are described.

1 Introduction

In today's computer systems databases have become very large and many tens of gigabytes are required to be on-line on the bigger systems. The analysis of a typical large transaction processing data base of a large government user showed that 67% of all disc transfers access a relatively small amount of data (less than 1 gigabyte). The performance of the disc system holding this data is critical to overall system performance since a very high traffic rate must be supported. The bulk of the remaining traffic is spread over many gigabytes of disc space holding "taxpayer" information; the discs holding this other data need to support a much lower traffic rate.

Fixed disc drives most commonly used for on-line data storage provide a compact method for storing large amounts of data but because of the mechanical delays access to data is relatively slow. Advances in semiconductor technology mean that it is now possible to produce "solid-state discs" of reasonable size at reasonable cost. The use of solid-state discs for the storage of frequently accessed data can improve overall system performance by reducing access time and so removing bottlenecks.

Many solid state disc systems have been developed both by manufacturers of mainframe systems and by specialist disc product suppliers. Often the design of the solid state disc system is specific to the system to which it connects (H. Sugiura et al., 1991); this is the case for ESS. The development of ESS was recognised as necessary to allow ICLs biggest customers to realise fully the performance potential of multi-node SX systems.

The design described in this paper uses state-of-the-art technology and incorporates a number of novel features.

2 Background to Requirements

2.1 Processor Performance

Between 1975 and 1990 the raw performance available from mainframe computers increased by a factor of about 25, being made possible by improvements in semi-conductor technology. Greater levels of integration on silicon resulted in reduced gate delays and hence lower processor cycle times; also as more gates per chip became available the functionality of the hardware could be greatly increased so that far fewer machine cycles per instruction were required. The time taken to access data held in various storage areas also had to be reduced to avoid this being a limiting factor.

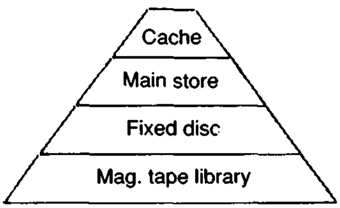


Fig. 1 Hierarchy of data storage in large computer system

The sizes of cache and main stores have increased while the access time has reduced to support improvements in processor performance. On secondary storage systems, now normally fixed discs, developments have taken a different direction.

2.2 Disc Performance

During the same period (1975 to 1990) the *capacity* of fixed disc systems have increased by a factor similar to the *increase in performance* of processors; in the mid 70's the typical capacity per disc spindle was 100 Megabytes; in 1990 2·5 Gigabytes was available. This increase was achieved by improvements to the materials used for the recording medium, the manufacturing methods and, also by changes to the technology of read/write heads (Isheida et al., 1991, Oshiki et al., 1991).

Over the same period there have been only small improvements in the performance of disc systems. The data transfer rate, governed by bit density and speed of rotation, has increased from around 1 MB/s to 3 MB/s. The access time, governed by the head movement and rotational latency, has improved by a factor of about 1·7, the average access time now being 24 milli-seconds. The divergence of performance between processors and disc systems has been recognized for some time, it has been referred to as the "I/O Access Gap" (Fujino et al., 1991).

2.3 *How to Close the Gap*

One approach to mitigate the effect of this gap is to reduce the number of disc transfers required by increasing the use of main store. The problems here are the cost of rewriting software both in operating systems and applications and, also a possible reduction in multi-programming ability of the system as main storage available for each application is reduced.

Another approach is artificially to reduce the capacity of disc drives by using only a limited number of tracks on a disc. In the limit, restricting use to just 1 track means that the head movement time can be eliminated completely; the access time is then only governed by the rotational latency. This approach is viable where the size of critical data is fairly small but becomes very unattractive from the point of view of cost and floor space as the disc space requirement increases.

The third approach is to eliminate mechanical delays in the disc system by the use of semiconductor technology – the so called "Solid State Disc". Of course this is not a disc at all but is an addition to the hierarchy of the system storage, hence the name "Extended System Storage" – "ESS".

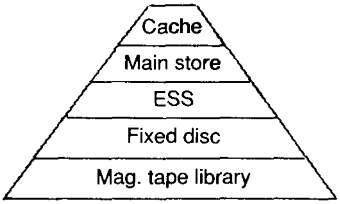


Fig. 2 Amended hierarchy of storage

3 Requirements for ESS

The main functional requirements for ESS agreed at the start of the project were as follows.

3.1 *Connection*

ESS must connect to a Series 39 Mainframe in the same way as a conventional fixed disc system. The reason for this requirement was to avoid major changes to the system architecture, the operating system and the hardware of the I/O system. Clearly there must be no changes required to any customer applications.

3.2 Performance

As shown earlier the seek time and rotational latency delays associated with conventional disc systems must be largely eliminated.

3.3 Integrity

ESS must exhibit the same characteristics of a conventional disc in terms of integrity of storage of data. The data must be non volatile, i.e. retained when mains power is removed; also sufficient error checks must be built in to ensure that undetected corruption of stored data is not possible.

3.4 Modularity

ESS should be modular so that its size can be matched to the needs of individual customers. The design should take account of future enhancements possible using next generation of memory chips.

4 Design of ESS to Satisfy the Requirements

4.1 Connection Method

To satisfy the first requirement for similarity to a conventional disc system, a number of design decisions were easy to make, namely the design must:

— use an existing disc controller
— use the same interface as an existing disc
— use the same addressing method as existing disc
— transfer data at the same rate as an existing disc

With these decisions the changes required in the controller would be restricted to firmware alone and would be mainly concerned with error management and reporting. The changes required to the operating system would be to add new unit descriptions for ESS. These recognize the differences in capacity from conventional discs and also allow the associated maintenance software (SAM) to handle ESS.

The high speed file controller (HSFC) shown in Figure 3 as the controller for conventional discs is based on a generic design HSPC (High speed peripheral controller) developed specifically for Series 39 and since used in various forms for connecting discs, magnetic tape systems, and various communications equipment. Connection of these various types of peripheral has been achieved by minor changes to hardware to accommodate different electrical interfaces and by writing alternative microcodes to support various command sets and protocols. For ESS it was decided to base the design of

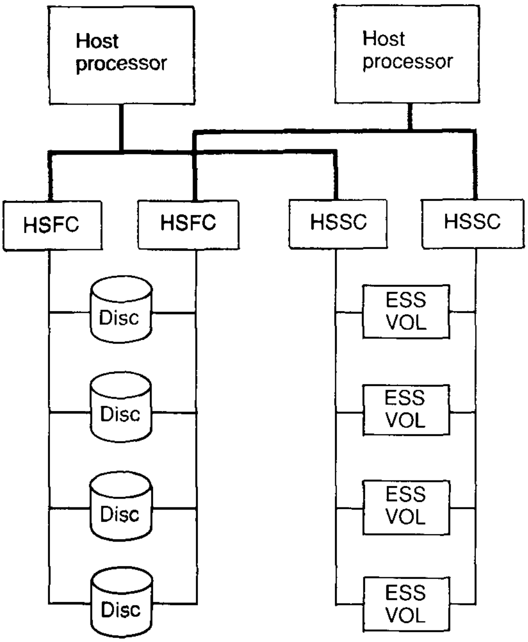


Fig. 3 System connection of discs and ESS

the HSSC (High Speed Solid State Controller) on the design of HSFC. By using exactly the same electrical interfaces any change to the controller hardware could be avoided. The interface used for ESS is the same as that used for an existing disc type – the Swallow Disc Drive supplied by Fujitsu. The use of this interface was seen as being low risk from the point of view of design complexity.

4.2 Performance

To eliminate seek time and rotational latency it is necessary to make the device entirely solid-state; the target was to make the total access time as seen at the interface to the controller less than 10 microseconds, which would improve average access time by 2500:1.

4.3 Integrity Against Power Failure

A key feature of conventional disc systems is their ability to retain data when the mains electrical power is removed. Clearly ESS will be relied upon in the same way so that its content must be retained for the duration of any power loss. The target agreed for the data retention time was greater than 24 hrs, on the basis that ESS will generally be used by customers with large non-stop operations most of whom have standby generators.

In some other manufacturers' products facilities are provided to store data on a local conventional disc drive, writing to which takes place when mains power is lost. Data is then read and rewritten to the solid-state store when power is re-established. This approach was considered for ESS but rejected on the grounds of cost and the risk to data integrity when the data is moved around without full system checks. A study showed that by careful design the 24 hr figure could be achieved using lead-acid batteries in the ESS cabinet.

4.4 Modularity and Capacity

Conventional disc systems currently used on ICL mainframes have capacities per spindle ranging from 300 megabytes up to 2.5 gigabytes. The larger discs have been introduced only recently and the majority of discs in use are either 640 or 1280 megabyte units. A design study showed that the module size for ESS should be 640 megabytes; this module would be called an ESS Volume. Up to 8 volumes would be fitted in a cabinet giving a total capacity of 5 Gigabytes.

5 Realisation of the Design

The ESS system is housed in two cabinets each 1441 mm × 705 mm × 805 mm. The controller cabinet contains a pair of HSSC controllers with their associated power system and optional CAFS units. The ESS store cabinet contains up to 8 ESS volumes and a power system including batteries.

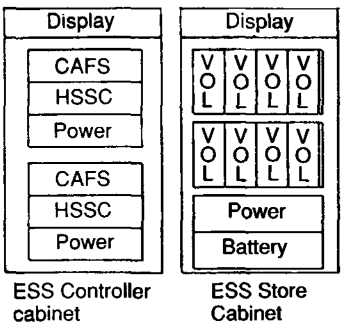


Fig. 4 Overall layout of ESS system

5.1 ESS Volumes

Each ESS volume has two main functional blocks; the store system and the interface control. The store system, comprising the printed circuit boards (pcbs) for the store, store control and error management control is designed as a package that can be reused with alternative interface controllers if required. The store system is designed for high performance, high reliability

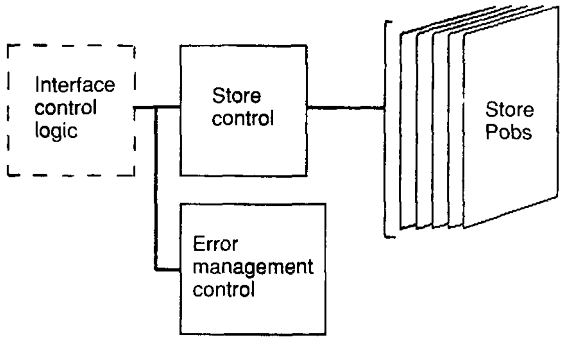


Fig. 5 Functional blocks of ESS volume

and low power consumption. The design features of the Store system units are now described.

5.2 The Store

The Store consists of 5 pcbs each containing 300 4-Megabit Dynamic Rams (DRAMS) arranged logically as 20 rows of 15 bits. 75 bits, held on 5 pcbs, are accessed in parallel; 64 bits are used for data, 1 bit for a special function, and 8 bits for error detection and correction using Hamming codes.

Data, address and control signals for the array of DRAMs are provided via a set of CMOS buffer devices mounted on the store pcbs. The DRAMs are also CMOS devices; use of this technology throughout means that power consumption can be kept to a minimum.

Use of CMOS drivers means that the consumption of an individual DRAM when not being accessed is 200 microamps compared with 2 milliamps had TTL drivers had been used.

The DRAMs used have an access time of 100 nS with a cycle time of 150 nS; taking account of the delays of drivers and interconnections, this means that the total bandwidth available at the store interface is approximately 40 Megabytes per second. Particular care was given to the allocation of driver components, to ensure that the failure of an individual driver will affect the operation of only a single bit of the store. As single-bit errors can be corrected this means that the system can continue to operate allowing the failing component to be changed when convenient.

The printed circuit boards allow double-sided surface-mounting of devices, the process for which was specially developed by the ICL Kidsgrove manufacturing plant for this project.

5.3 The Store Control

The Store Controller is a custom-designed VLSI chip contained in a 256 pin package. The technology used is Bi-CMOS which gives the benefits of high speed, acceptable packing density, and low power consumption. The main functions of the store control chip are to provide the data, address, and control signals to the store pcbs, to handle detection, correction and reporting of store errors, to provide the refresh function for the DRAMs and to provide a standard interface to the rest of the system.

The store control chip receives addresses and data via a 32-bit interface. For write cycles the data is first passed through a Hamming code generator; the code bits are appended to the data which is then passed to the store with the control signals necessary to perform a write function. In the case of a read cycle the store address is passed to the store with the control signals set to perform a read; the data is received in the store control and is passed through Hamming checking logic; in the event of a single bit failure being detected this is corrected before the data is passed back to the requesting unit. Where the request originated from the error management chip the store control chip has a special function to perform. Where a single bit failure is detected the failing bit is corrected and the resulting data rewritten to the store. The address is then re-read and the Hamming check performed again; if the failure is still present it is logged and reported to the system as a "hard" fail (eventually requiring maintenance action). In most cases the failure will have been a transient or "soft" failure and will have been corrected by the action of rewriting the data. The soft failure rate of DRAMs could cause a problem if this approach was not adopted as it would be possible for single soft fails to accumulate in the store until eventually two single bit fails combined to cause a non-recoverable multi-bit failure.

The store control chip was designed to support 2 refresh rates, one cycle per 15 microseconds or one cycle per 128 microseconds. The advantage of the slower rate is that it conserves power by activating logic in the DRAMs less frequently. The faster rate was included because of uncertainty, at the start of the project, about the supply of DRAMs with a capability for extended refresh timing.

At the start of the project a decision was required on the size of DRAM to be used; since the introduction of the first DRAMs in 1970 the size of device available has increased by a factor 4 every three years. When a new generation of DRAM is introduced its price is initially very high. As production volume and sales increase the price falls until at some point it becomes cheaper in terms of cost per bit than the previous generation. Historically the time from introduction to "price per bit cross-over" with the previous generation has tended to be about the same, although subject to some variation for various reasons. For the most cost-effective solution a number of options in the design allowed a final choice to be made later. The development and production lifetime of ESS would embrace use of 1 megabit, 4 Megabit and 16 Megabit generations; it was decided therefore that the design should cater for all three types. The type actually used would be indicated to the store control chip by a code provided by the store pcbs; the store control chip provides alternative address decoding depending on the type used.

5.4 *The Error Management Control*

Error management control is implemented in a custom chip using the same technology as the Store Control. The error management chip connects to the store control chip via a bus that is shared with the interface control; there is also a direct connection to the interface control continually monitoring whether the ESS volume is being accessed by the system. When accesses from the system are in progress the error management chip does nothing except monitor the error lines from the store control. When the system is not using the ESS volume, the error management chip performs a function called "trickle read", which involves instigating read cycles by the store control at successive addresses through the store. The trickle read function working with the error detection/correction/rewrite sequence described earlier keeps the store clear of "soft" failures whilst "hard" failures are reported to the system. If no system accesses are in progress the entire store can be checked in less than 2 minutes. The other main function of the error management is to provide a store test and initialisation ability, for use by an engineer after repair or at the time of initial installation of the system.

5.5 *Store System Parameters*

The design of the store, store control and error management took into account possible future requirements to support alternative interfaces. The number of interface control signals has been minimised and the timing constraints and protocol kept simple. The interface provides a bandwidth in excess of 10 Megabytes per second. Power dissipation has been minimised by careful choice of components and when not being accessed is only 5·5 watts for a full volume. The error correction and trickle read features give a high reliability design; the mean time between failures causing loss of data is estimated to be in excess of 100,000 hours.

5.6 *Interface Control Logic*

5.6.1 Serdes (Serialiser/Deserialiser) The function of the Serdes is to convert the bit serial data stream received from the controller into byte wide data for the interface control chip (ICC). The reverse function is performed for data read from the ICC to be passed to the controller. The design is implemented in a high speed ECL VLSI chip already in use on another project.

5.6.2 Control Decode The control decode logic is implemented in re-programmable devices (EPLDs) to allow minor changes to the command

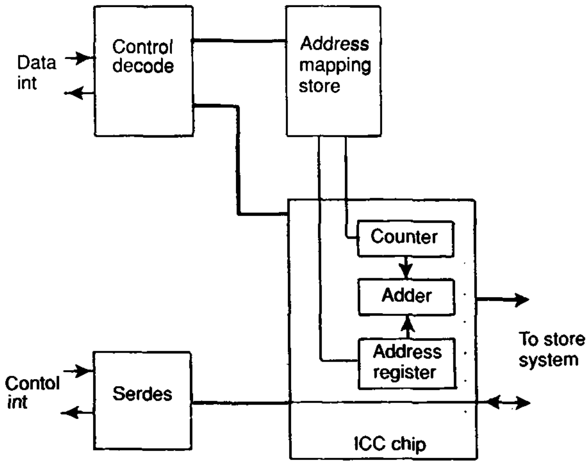


Fig. 6 Shows a block diagram of the Interface Control Logic

and response codes used by the controller. The main function of this logic is to receive commands and parameters from the controller and to pass status and error information back to the controller.

5.6.3 Interface Control Chip and Address Mapping Store The ICC is the piece of design that emulates the operation of a conventional disc; to explain its operation it is first necessary to understand addressing of conventional discs.

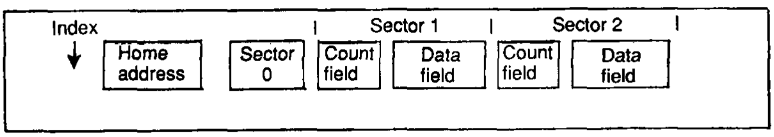


Fig. 7 Shows the format of data stored on a track

Disc Addressing: The start point of a track is known as "Index". The first 2 data fields on the track are known as Home Address and Sector 0; these contain information about the track and are for use by the operating system. The remainder of the track is available to a user who has control over the size of data block. A track on a disc is selected by a "cylinder address" and a "head number" sent by the controller. The data block required can then be found either by reading along the track or by a method called rotational position sensing (RPS). RPS involves the disc receiving a position on a track from the operating system and then finding that position without any interaction with the controller.

ESS has to support exactly the same addressing mechanisms as outlined above. The equivalents of disc tracks with home addresses, count fields, data blocks and gaps are written to the DRAM store. The main function of the ICC is to convert conventional disc addresses to store addresses. It must also emulate alternative disc geometries. The table below indicates the disc types emulated. It should be noted that because ESS has a fixed capacity it is not possible to emulate the full capacity of all types of disc. Emulation in all cases is of the cylinder size of the equivalent disc type; a whole number of cylinders is always emulated. The address mapping store is the method used to handle these alternative emulations.

Disc Type	Number of Heads	Cylinder Capacity	Number of Cylinders
FDS 2500	15	745920 bytes	899
FDS 5000	23	1143744 bytes	586
FDS 20G	19	944832 bytes	710

The address mapping store consists of three Read Only Memories (ROM); two are used to translate head and cylinder addresses to store addresses, the third to translate RPS values. The ICC contains an address register, an adder, and a counter which together generate store addresses as follows:

> A head and cylinder address from the controller is used to address the head and cylinder translation ROM; the data read from the ROM is loaded into the ICC address register and represents the start address of disc track (equivalent to Index). In the case of a track search the ICC counter starts counting from zero, the output of the counter being added to contents of the address register. The counter increments once every 50nS; this is the same as the bit rate of the data passed across the interface to the controller. Every 64 counts a store cycle is initiated.

In the case where RPS is used the RPS value is used to address the relevant ROM; the contents read out are used to pre-load the counter with a value representing the appropriate offset along a track. The method of synchronizing the data and finding the appropriate count field is similar to that employed for a normal disc. The address mark is stored as the special function bit in the store; when the ICC reads this bit from store it communicates an "address mark found" signal to the controller which then acts as it would for a normal disc.

5.7 *Design for Data Integrity*

The data integrity of the system is catered for at three levels, viz *operating system and controller checks, interface logic checks* and the *store system checks* already described.

The operating system and controller work together to perform a check on the content of count fields before a read or write of a data block starts. This check ensures that the block accessed is indeed the one addressed. A lower level check on the addressing mechanism is implemented in the hardware of the interface control; the content of the address mapping stores are parity checked and a further check is provided on the counter and adder in the ICC. These low level checks are particularly important during the process of formatting the track during which no operating system or controller checks are possible. The interface between the ICC and the store system has parity checks for data, address and, control codes.

5.8 Design for Minimum Power Consumption

To maintain stored data during long periods without mains power special measures had to be taken. One approach was to use components with low power dissipation and so to combine these components as to minimise power consumption as in the store pcb design described earlier. Another consideration was the minimum set of components to be "powered on", to retain stored data; the remaining components could then be switched off when mains power is lost. As a result the power supply for the store system and the power supply for the interface logic are separated. The power system provides and controls these separate supplies.

5.9 ESS Power System

A simplified schematic of the power system is shown in Figure 8.

Clearly the power system is fundamental to reliable retention of data and the design must be exceptionally reliable. The approach has been to duplicate

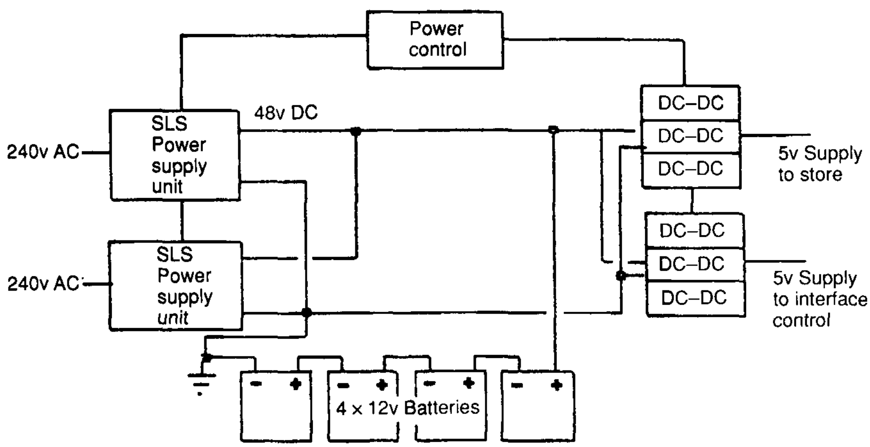


Fig. 8 ESS power system

most of the major components and, where this is not possible, to use components with very high intrinsic reliability.

The operation of the power system is as follows:

> The 48 V DC supply provided by the two SLS power units is used to charge the battery system and to drive the two sets of 48 V to 5 V converters. As described above the two sets of converters provide separate power to the store and interface circuits. As can be seen the converters are arranged in 2 groups of 3; in both groups, only 2 converters are required to supply the power so the failure of one in a group will not affect system operation. When mains power is lost the batteries continue to provide the 48 volts to the converters. The various power supplies, converters and batteries are monitored by the DC and Environmental Control.

5.10 DC and Environmental Control

This is a microprocessor based unit which performs various monitoring, control and, reporting functions concerned with the power system and cabinet environment. In summary the functions supported are:

- Power supply control, switching on/off, monitoring output
- Battery control, maintaining on float charge, monitoring condition during mains power off, estimating battery life
- Monitoring cabinet temperature conditions
- Monitoring and controlling fan speed
- Reporting failures and warning conditions to the system via the interface control

The DC and Environmental control operates from its own auxiliary power supply unit supplied by either mains or by the battery.

6 Design and Product Validation

The design validation of ESS concentrated on ensuring that the VLSI chip designs were correct to avoid expensive iterations. The approach was to use a top down methodology; from specification to high level model followed by capture and gate level simulation.

After prototype build and switch on the process was one of proving the design as an independent unit and then working with other system components. The testing goes to great lengths to prove all the functionality and particularly the error management system. The overall reliability of the system is closely monitored during the validation process to ensure that the predicted reliability is achieved. A number of minor design problems were found during the testing of prototype units but it was possible to overcome these without any chip iterations.

The final stage of the validation process is the testing to ensure that the product meets relevant ICL and international standards; this includes environmental testing (temperature, shock, vibration etc) and also EMC and RFI testing.

7 Summary

The justification for the development of ESS was mainly on the basis of the improvement it would give to system performance. Although the actual improvement on any one system cannot be accurately predicted, the measurements taken as a part of the validation process confirm that the performance of the ESS device are in line with expectations and will give large improvements in the number of transfers per second that can be supported. Figure 9 demonstrates the improvement for some typical cases.

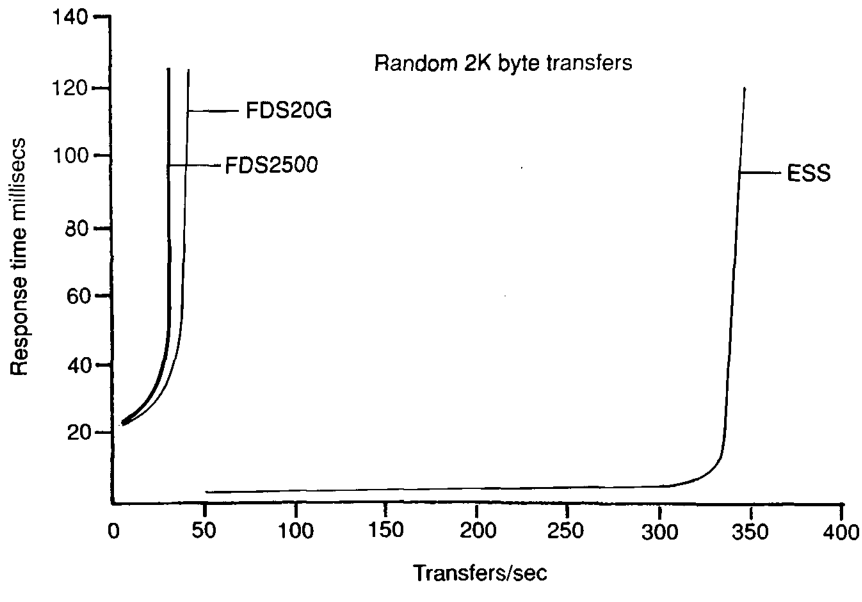


Fig. 9 Disc and ESS performance

Users of ESS will be able fully to exploit the performance available from their large systems by removing disc system bottlenecks. This will be achieved by the user moving the most frequently accessed files from conventional disc to ESS, the modularity, size and compatibility with conventional disc types will all contribute to making this a straightforward process.

References

SUGIURA, H. et al.: F6631 Solid State Disk: High Speed Virtual Disc Unit. *Fujitsu Scientific & Technical Journal* (Feb 1991).

ISHIDA, S. et al.: Thin Film Disk Technology. *Fujitsu Scientific & Technical Journal* (Feb 1991).

OSHIKI, M. et al.: Thin Film Head Technology. *Fujitsu Scientific & Technical Journal* (Feb 1991).

FUJINO, M. et al.: File Technology Overview. *Fujitsu Scientific & Technical Journal* (Feb 1991).

Biography

A.W. Jenkins

Tony Jenkins joined ICL in 1965 with an HNC in Electrical Engineering from Oldham College of Technology. He has worked exclusively on hardware development projects specialising in memory system design but with significant involvement in the design of two versions of the Distributed Array Processor. He is currently working in the Systems Engineering Group of Corporate Systems Division where he has had management and overall design responsibility for ESS.

ICL TECHNICAL JOURNAL

Guidance for Authors

1. CONTENT

The *ICL Technical Journal* has a large international circulation. It publishes papers of high standard having some relevance to ICL's business, 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 applications must include some suitable introductory material or reference.

The Journal 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 reviews may also be published.

2. AUTHORS

Within the framework defined by §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. ABSTRACTS

All papers should have an Abstract of not more than 200 words, suitable for the various abstracting journals to use without alteration. The Editor will arrange for each Abstract to be translated into French and German, for publication together with the English original.

5. PRESENTATION

5.1 Printed (typed) copy

Two copies of the manuscript, typed $1\frac{1}{2}/2$ spaced on one side only of A4 paper, with right and left margins of at least 2.5 cms, 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 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.

5.3 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. Short tables can be given in the text but long tables are better submitted on separate sheets and these, as for diagrams, must carry the author's name and the title of the paper.

5.4 Photographs

Black-and-white photographs can be reproduced provided they are of good enough quality; they should be included only very sparingly. Colour reproduction involves an extra and expensive process and will be agreed to only exceptionally.

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, and only text references; 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 Journal's 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 25 offprints of their papers, free of charge, and further copies can be purchased; an order form for copies is sent with the proofs.

8. COPYRIGHT
Copyright in papers published in the *ICL Technical Journal* 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.