2-1-1998

A general model for print delivery of internet documents

Mohan Dhandapani

School of Printing Management and Sciences

Rochester Institute of Technology

Rochester, New York

Certificate of Approval

Master's Thesis

This is to certify that the Master's Thesis of

Mohan Kumar Dhandapani
name of student

with a major in Graphic Arts Publishing

has been approved by the Thesis Committee as satisfactory

for the thesis requirement for the Master of Science degree

at the convocation of

February 1998
date

Thesis Committee:

Frank Cost
Thesis Advisor

Marie Freckleton
Graduate Program Coordinator

C. Harold Gaffin
Director or Designate

A general model for print delivery of Internet documents

by

Mohan Kumar Dhandapani

A thesis project submitted in partial fulfillment of the requirements for the degree of Master of Science in the School of Printing Management and Sciences in the College of Imaging Arts and Sciences of the Rochester Institute of Technology

February 1998

Thesis Advisor: Professor Frank Cost

Permission to Reproduce the Thesis

Title of Thesis

A general model of print delivery of Internet documents

I, Mohan Kumar Dhandapani, hereby grant permission to the Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: 2/23/98

Acknowledgments

I would like to thank my family for giving me the moral support to make this all possible. I am greatfull to Frank Romano who has been a source of inspiration and had supported me in various ways during crucial time. I am also thankful to Dr P R Mukund, at Electrical Engineering department, who has been an indirect help in fulfilling my dream. My sincere thanks to Michael Robertson, at the Wallace Memorial Library, for guiding me at various stages of this project and also for supporting me towards my graduation.

I would like to thank Marie Frackleton, my course coordinator, for guiding me through out my graduate program and for the support she extended in preparing my thesis materials.

I would like to thank my thesis advisor, Frank Cost, who's vast knowledge in this subject and the time he spend on the work has made this thesis work a valuable resource.

I would also like to extend my thanks to Barbara Birkett for the assistance extended to experiment with the Docutech System.

A special thanks to Pat Pitkin, Director of Wallace Memorial Library and all the staff of systems department who gave me timely help on this project.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ASCII	American Standard Code for Information Interchange
CCC	Copyright Clearance Center
CD-ROM	Compact Disc - Read Only Memory
CDF	Channel Definition Format
CMYK	Cyan Meganta Yellow and Black
CPU	Central Processing Unit
DCS	Desktop Color Separation
DCT	Discrete Cosine Transform
DigitalID	Digital Identification
DIP	Desktop Internet Publishing
DPC	Desktop Publishing Center
DTP	Desktop Publishing
EPS	Encapsulated PostScript
FTP	File Transfer Protocol
GIF	Graphic Interchange Format
HTML	Hyper Text Markup Language
HTTP	Hyper Text Transfer Protocol
IP	Internet protocol
ISDN	Integrated Standardized Digital Network
JFIF	Jpeg File Interchange Format
JPEG	Joint Photographic Experts Group
LAN	Local Area Network

MB	Mega Bytes
MIME	Multipurpose Internet Mail Extension
OPI	Open Prepress Interface
PDF	Portable Document Format
PS	PostScript
RAID	Redundant Array of Independent Drives
RIT	Rochester Institute of Technology
SGML	Standardized Generic Markup Language
SPMS	School of Printing Management and Science
TCP/IP	Transfer Control Protocol / Internet Protocol
TIFF	Tagged Image File Format
URL	Universal Resource Locator
WML	Wallace Memorial Library
WWW	World Wide Web
XML	eXtended MArkup Language

Abstract

With the explosion of the Internet there are abundant opportunities for budding authors (writers and designers). Their content can be eaily posted on the Web and accessed by a wide reader base through WWW. However, this on-demand and on-site publishing is limited to on-screen viewing and desktop printing. As more and more books and materials are accessed, there is growing need for on-demand printed copies.

Ordering a printed copy is still a traditional business which limits amature authors and users from getting easy access to them. Some of the on-line bookstores provide only the purchase transactions through on-line, while the printing itself is done through traditional process. Some of the growing needs of on-demand printing include:

- getting a printed copy of a electronic Thesis material,
- printed copy of selective sections of a User Manual,
- revised pages of a Book.

This thesis project involves a thorough study of a Model to facilitate on-demand print of documents available in Internet covering such issues like quality, speed, copyright, security, bandwidth, royalty and delivery. A working project will be developed, demonstrating the Model, using a Docutech Printer. The steps involved in setting up a work flow to facilitate on-demand printing of an Internet document using Interdoc/Docutech work flow will be documented.

This work could be further extended to adopt to the evolving "Collaborated Publishing" concept widely being discussed for use in the Academic Society.

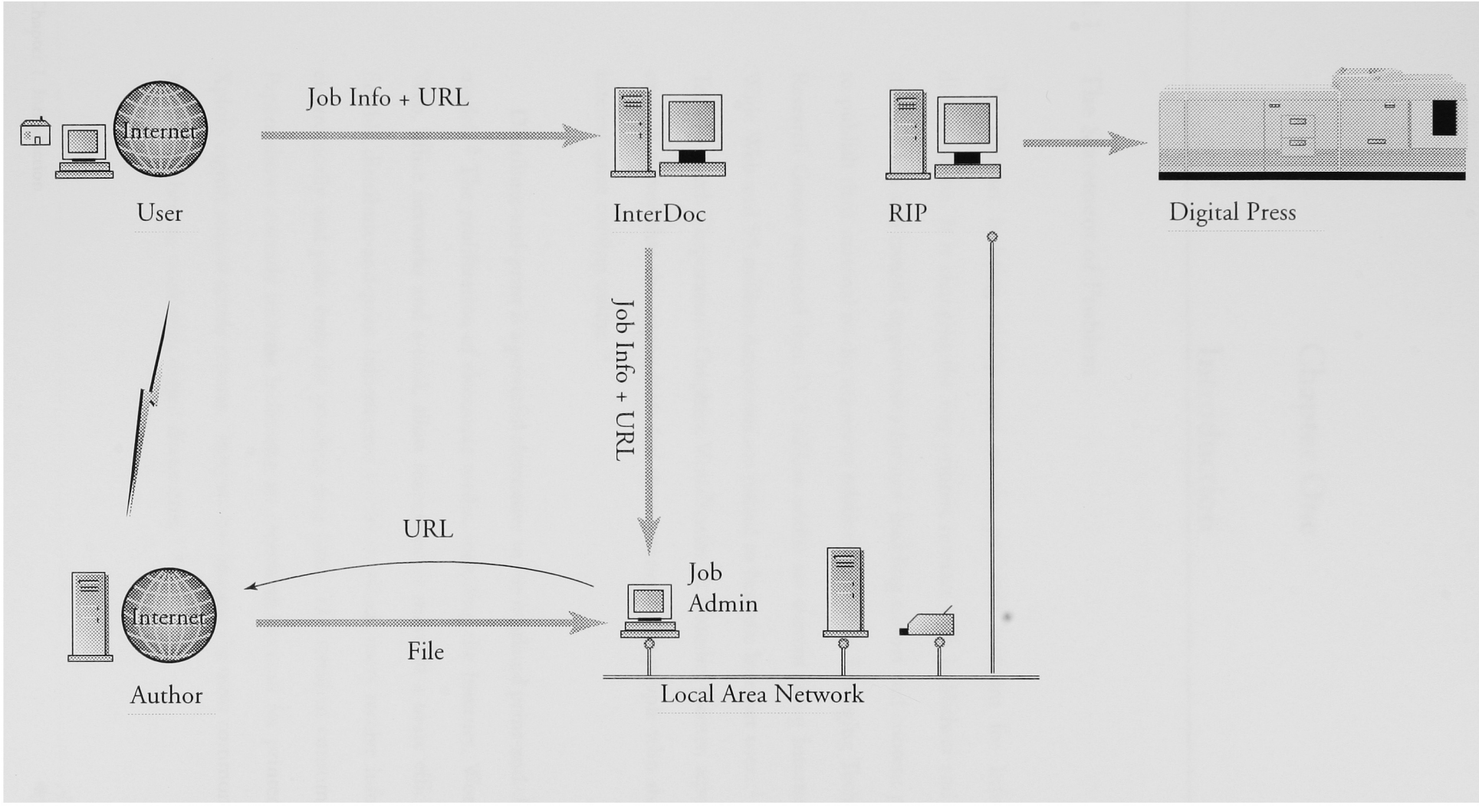


Figure i, Flow Diagram proposal for alternative print delivery of Internet documents

Chapter One

Introduction

1.1 The Statement of Problem

The Internet is being widely used as the preferred medium for Information dissemination. It is changing the way printers interact with publishers and the end users. It has also created opportunity for new budding editors and content providers to publish their material to the masses at relatively low cost. Emerging Technologies Research Group reported that 31.3 million adults are current using Internet World Wide Web and 55 million Americans are poised to become Internet users.[1] Georgia Tech Research Corporation's Graphics, Visualization & Usability Center, reports in its seventh survey that the respondents feel more connected to people who share their interest since coming online. [2]

Distribute-and-print is a powerful alternative to the outdated print-and-distribute model. [3] The proliferation of electronic media, including the Internet, World Wide Web, office networks and e-mail, allow businesses to move to a more efficient and flexible distribute-and-print information model in which users receive information electronically and print only the portions they need. The trend in consumption of Papers shows a steady increase indicating an increasing demand for printed copies. Xplor's eighth annual sttudy reports. "Internet has become the most common way for Xplor members to work with digital documents." [4]

Such explosion in information brings to light some interesting problems.

- The content providers would like to use the Internet as an advertisement medium to get browser's attention while they look for a standard model to charge them for getting the full content.

- On-line books are currently an afterthought of printed copies. Ideally, the content should be generated in a generic format suitable for both on-line and print on-demand dissemination.

- The issues of copyright and royalty are subjected to government and geographic regulations. Authors and browsers are forced to go through the traditional publishing bottleneck to handle such regulations.

- Although it is easy to display information retrieved from the Internet on a monitor, it is not easy to print that information in the format in which it was originally created.

1.2 Background and Significance

We have entered the era of Information Age. Today, business, research, school, government, military, entertainment and almost all the life line functions depend on information for their achievement. Timely information is proving to be more and more valuable. The Internet is providing the best media to facilitate information

sharing. It hs enabled geographical cross over and is bringing past back to life and helps redefine our future in a more cooperative and collaborative way.

Historically, with Gutenberg's gift of the movable type, a need to disseminate more information was perceived. Publishing evolved as a trade, linking the content owner and the reader. With todays' Internet explosion, publishers' expertise will be a great asset in regulating the information jungle.

Internet, is it a threat or a boon? Is it misguiding our teens instead of empowering them to face the future? Where do you lay the boundary between the copyright and marketing? Who polices the information? These are the hot topics of todays' debate. Ironically, publishers and printers who have centuries of experience in information dissemination are keeping a very low profile in shaping the future of the information industry.

Researchers and news agencies are among the top users of the true power of Internet. The Internet is attracting more information than any one could predict. This shared database resource is widely being used and being built upon. We thus have entered the collaborative publishing era.

Enabling the user to easily print the information that is as context-rich as it appeared on the screen will greatly advance the distribute-and-print trend that is currently being adopted by businesses and homes around the world. As more and more people embrace the distribute-and-print trend, they will discover improved productivity, greater convenience and increased control over the information they receive.

1.3 Reasons for Interest

The future of information dissemination has been envisioned to be completely different than what we have today. Collabrative publishing, on-demand customized information, constantly evolving standards, cross boundary sources, speed of retrieval and a host of new concepts are creating a turmoil in the information industry.

This is not a wait and see game. There is no predictable end. Publishers with their rich tradition and experience have to get on the bandwagon, if it is not already too late. It is encouraging that the Internet explosion constantly creates opportunity for publishers. A need to organize information, moderate retrieval and sharing, use available media to communicate to the wide reader base, satisfy legal and global requirements and the like.

Knowledge of publishing, printing, Internet backbone and a vision of the future is required to point the proper direction for publishers and enterprising printers.

Endnotes for Chapter 1

1 Miller, Thomas E. *The 1997 American Internet User Survey.* Emerging Technologies Research Group, 1997.
World Wide Web site http://etrg.findsvp.com.

2 *GVU's 7th WWW User Survey.* Graphics, Visualization & Usability Center, Georgia Tech Research Corporation, 1997.
World Wide Web site http://www.cc.gatech.edu/gvu/.

3 McKenzie, Matt and Votsch, Victor. *Publishing Alternatives to the Web.* The Seybold Report on Internet Publishing, Vol 1, No 3, pp. 3-7. Seybold Publications, 1996.

4 *Internet is making a big impact on high-volume printing.* The Seybold Report on Internet Publishing, Vol 1, No 3, pp. 32. Seybold Publications, 1996.

Chapter Two

Background Theory

The Internet is undoubtedly the single tool revolutionizing almost all major industries. The most dramatic of all is the information industry. Vast advancements in network, telecommunication and document standards enable new growth opportunity in information internetwork. Almost all major computer related vendors are working on Internet related products and are releasing advanced products at an ever increasing rate.

Electronic publishing products are switching from traditional publishing to Internet based publishing relatively easier.[1] Consumer software products and the regular office Suites are packed with full blown features for Internet publishing and communication.

Hardware improvements and mass production has brought affordable power to desktop users. The Internet Link is almost a standard feature with any PC today. Abundant freeware and shareware are making it even more easier to learn the new trade. Amateur authors are are springing up from all around the globe.

This thesis project is an attempt to research the current available technology and demand, look into the future of the information dissemination process and develop a model for print delivery of on-line electronic documents.

2.1 Digital Printing Press

At the IPEX-94 trade show, Indigo introduced its full color digital press — Indigo E–Print 1000. Since then, there has been tremendous technological development in the digital press field. The products range from the desktop printers, networked groupware printers to fully computerized high–speed printing presses.

Products like the Xerox DocuTech 6135, Xerox DocuColor 40, Xeikon DCP-1, Agfa Chromopress, IBM InfoColor 70, Indigo E–Print 1000 and T/R System's MicroPress 312 has proved the feasibility of on-demand and variable data printing. On-demand printing is the fastest growing segment of full-color printing. Reduction in makeready time, consistent quality from single print quantity to thousands of copies, rapid change of jobs on the press, economical short-run print, meeting "Just-In-Time" demands of the customers are all favoring the trend towards the digital press.

All the digital presses in the market are able to accept industry standard PostScript documents; thus, enabling customers to distribute their electronic files to their desired destination and print them locally saving shipment cost and time. Feasibility to work from digital data also brings advantages like ability to include last minute information.

The high investment cost of these machines prevent its reach by small and medium size printers. Need for a rapid change in their work flow to suit the digital press also makes transition to digital press slow.

In spite of the hurdles posed by to day's technology, the prediction is clear for a rapid movement towards short and on-demand customized printing.

2.2 Internet Architecture

The Internet is defined by technologist as a network of networked computers. From the user end, Internet is a system for instantaneous distribution of digital information from the source computer to any number of linked user terminals at a relatively low cost.

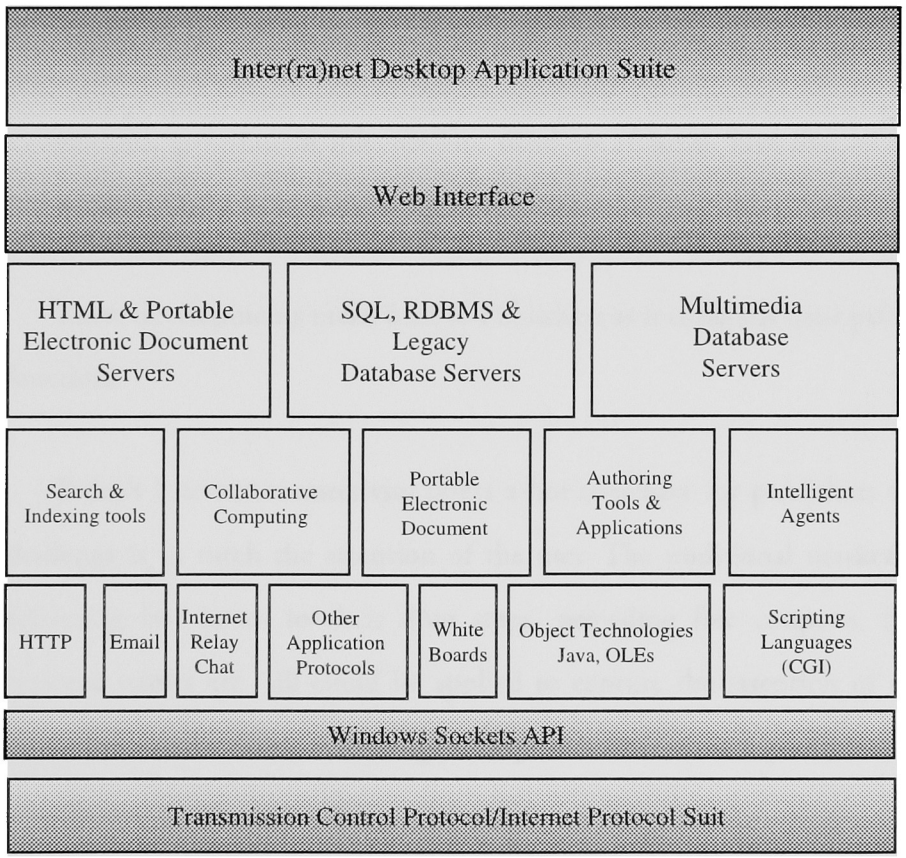


Figure 2.1, Internet Topology

Internet extends the computer–to–computer network widely used in the UNIX computing arena to any TCP/IP compatible network product. The standards of TCP/IP connection which was originally developed to enable cross computing between government and research organizations became open for commercial use. With the demand ever increasing due to the thirst for information and connectivity, the products became increasingly cheaper. With more players in the market, Internet products are as diverse as any other software in the industry today.

Internet features include static text, computer graphics, true audio, true video, virtual reality and the possibility of smell and touch in the future. Internet applications include on–line shopping, news update, on–line publishing, financial transactions, government information, database archival, Live link to space stations and projects and a host of other opportunities.

These developments mean a lot to Publishers as it concerns their primary business function.

Today's Internet connectivity offers a hot user base for publishers to reach. The challenge is to catch the attention of the user. The traditional marketing tactics of delivering free issues to their door steps, providing free coupons, printing their personal names etc.. all could be applied to capture the attention of these readers. Better yet, the following technological improvements make it possible to invent more creative marketing strategies.

2.2.1 Personal Email

Email is the ignitor of the Internet explosion. The possibility that two person thousands of miles apart can communicate almost instantaneously out such low cost, bundled with the possibility of broadcasting, attracts people from all walks of life.

Electronic mail is an extension of the UNIX built-in mail service. It was available for a long time before it was commercialized. As the demand increased, dedicated email servers were installed (and are being installed) at relatively low cost and effort all around the world.

This communication medium is widely being used to transfer digital files across the globe. Information including pictures and application files are communicated with relative ease and low cost. This facilitates distribute–and–print concept and brings on-demand and localized printing a possible alternative to the mass print and distribute concept.

2.2.2 Push over Pull technology in WWW Publishing

So far the Internet has been viewed as an information source where the content provider dumps the material and the reader searches through to PULL the necessary information.

With the PUSH concept, similar to subscribed magazines, the information is PUSHED to the user by the content provider (or news filtering agencies). This opens up a better opportunity to define a niche market based on customer interest.

2.2.3 Selective Updates

The same complaint about newspapers not being fully read by any single reader applies to Internet information. While the user patiently downloads the information, it is not fully read. Technological developments are made to prevent this waste of bandwidth and to deliver only the requested material and to intelligently determine if the information is already available on the user desktop.

2.2.4 Digital IDs

With the issue of privacy and profitability becoming an increasing issue, developments are taking place in the area of identifying the content provider and the user through a digital ID. This establishes a personal connection between the two which can be used for security purpose or for commercial charging purpose.

2.2.5 Active Desktop

The next wave is to make the desktop of the computers take over most of the frustrating and time consuming search/retrieval/selective updating work of the user. With NetPC (networked personal computers) not very far away, active desktop will become an essential aspect of the daily interaction wit the information jungle.

2.2.6 Internet Web Television

Bringing the convenience of television and functionality of Internet is the next leap in information dissemination. Television will no longer be a passive broadcast receiver, it

will be active and will be able to fetch what each family would like to have from the huge information mess.

The technology is already there while it is a long wait to see an attitude change in the use of television as an interactive equipment.

2.3 Networked Printing

When centralized computing used to be the norm in the computing world, centralized printing was widely used. All the printing needs of a corporate would be routed to a network printer in the network.

The trend changed with the introduction of economical Desktop printers along with the power of desktop personal computers. While the concept of shared data was still practiced, the applications were not shared between desktop units and the main server. Network licensed software products became very popular.

With the growth in Internet, the concept of powerful central servers and networked printing is getting back to its day. Over the Internet, the documents origin could be any where on the global network. Potential readers who need to get a copy of the document uses their Desktop printer to print a hard copy. Corporate companies are turning to networked printing to provide fast and added finishing features for such printing.[2]

Corporate companies are moving towards groupware products.[3] These enable document network along with the project/production workflow. Internet based Print

administration products enable routing documents to a remote printer through Internet backbone.

2.4 Web Server Technology

World Wide Web Servers are becoming increasingly common and easy to install and maintain. Practically any Unix/Windows NT and some Mac Servers could be hooked onto the Internet. An Internet Protocol (IP) Address is required to identify each server on the network.

A Web server in its simplicity is a file server adopting the HTTP (Hyper Text Transfer Protocol) standards to communicate with the clients and other servers on the Internet. Files are of type text/binary with certain MIME types like HTML, Plain text. GIF images. JPEG images and the like can be transferred between a server and a client.

Materials over the Internet are viewed by the end user through a Web Browser application. The power of the Web is derived from two sources: HTTP and HTML. HTTP is a lightweight stateless networking protocol that uses minimal network bandwidth. In addition, its simplicity makes it easy to design and implement an HTML server or client (browser).

More advanced servers are able to handle other file transfer formats like Emails, Bulletin board, Chat and shared application. Figure 2.1 illustrates the gamut of an INTERNET setup. Almost all corporate business functions are integrated.

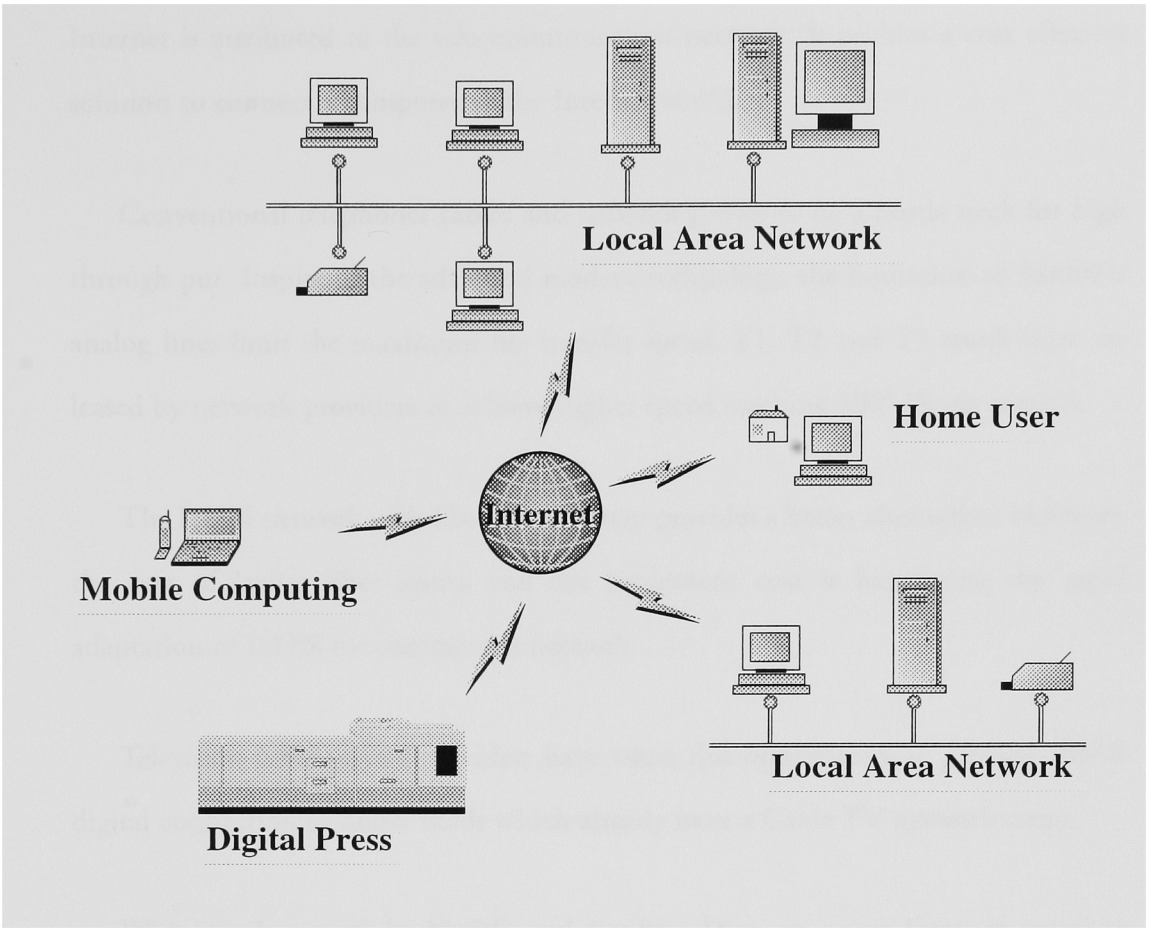


Figure 2.2, Internet Lan Network

Based on the building blocks, services like searching, collaborative computing, digital document transfer and other intelligent scripting applications can be developed. Advanced applications include merging existing database resources and serving multimedia and live information on the fly.

2.5 Telecommunication Network

Today, the backbone of the Internet is the telephone network. All commercial routing of the Internet data is through conventional telephone lines. The rapid growth of

Internet is attributed to the telecommunication network. It enables a cost effective solution to connect a computer to the Internet world.

Conventional telephones cables and network proves to be a bottle neck for high through put. Inspite of the advanced modem technology, the limitation of 3,000Hz analog lines limit the maximum file transfer speed. T1, T2 and T3 trunk lines are leased by network providers to achieve higher speed reaching 100MB per second.

The ISDN network with fiber optic system provides a better alternative. However, the cost of laying fiber optics and the equipment cost is hampering the rapid adaptation of ISDN for commercial network.

Television cable service providers have taken this opportunity to provide 10MB digital connection to house holds which already have a Cable TV network setup.

With developments in NetPC and Satellite Disc, one way Internet access is provided is through a satellite network. This enables Internet connection to remote places and for mobile computing.

2.6 Digital File Transfer

HTTP format is the most common file transfer format in the Internet. HTTP file transfer is based on MIME types and is suitable for data files like HTML documents and standard images.

File Transfer Protocol (FTP) is widely used which facilitates file transfer between two different computer platforms to send and receive files using Transfer Control Protocol/Internet Protocol (TCP/IP) network stacks.

This cross platform network proves very useful in establishing links between clients and service providers. In the graphic industry, PostScript and PDF files are becoming the standard file formats. These files formats are device and platform independent and hence are highly suitable for FTPing between different computing platforms.

Sharewares and freewares are available over Internet which makes FTP as easy as moving files in the local storage. Microsoft Office 97 suite, BBEdit, Netscape and almost all new office suits provide built-in FTP features to communicate with remote Internet hosts.

2.7 Archival and Retrieval

Networked computing and distributed resources is bringing central file servers back to life. More and more companies are using juke boxes and RAID hard disks to store all the documents. Network management tools enable proper routing of these documents with time stamps and progress marks.

Central administration of documents in a generic format helps in efficient archival and retrieval systems. Also, on–the–fly conversion of data to desired document format helps organize data for repurposing.

Corporate companies are able to make use of their central data resource for networked printing and variable data printing. Powerful search engines like Verity and Cascade are able to index documents of varies formats including PDF. Image archival systems based on Image content and keywords are providing rich search and retrieval system for Image database as well.

2.8 Electronic Publishing Market

Electronic publishing is being revolutionized by computer technology based companies. The traditional print media publishing has just started to feel the impact of the technology — or rather a paradigm—shift in the age–old practice of publishing.

A careful study and close monitoring of the following markets would reveal best the trend in electronic publishing as it stands today.

2.8.1 News Paper

Mass media news dissemination is the all major purpose of publishing. With the thirst for information, news agencies and news media are ever expanding and adopt to the latest technology in publishing and broadcasting.

2.8.2 Digital Libraries

Universities with funds from government and other research funding are the next best places to watch the trends unfold. They have the expertise, the need to share

information and the non-commercial form of expense budget! to develop new technologies.

2.8.4 Corporate Publishing

The need for commercialization is best understood in today's competitive corporate world. The huge information data banks, the speed at which updates are required, the shorter and shorter time–to–market and the global position of most these companies force them to look out and exploit the best technology for information publishing/ distribution.

2.9 Document Formats

2.9.1 Digital Document Formats

While all digital information today is a basic binary file, and that TCP/IP provides facility to transfer between cross platform, there are unlimited number of file formats. Digital images serve a multitude of purposes. Archive, compression, network transfer, manipulation, enhancement are some of the primary purpose for digitizing images. The wide range of originals, input/output devices, applications and publishing media demand specific image format for optimum reproduction. Each format differs in how it packs and structures the data. A simple text file is based on a 7–bit American Standard Code for Information Interchange (ASCII) format. Adobe's PostScript is the most common file format in graphic industry. PDF is another file format which is an restructured PS file. Tagged Image File Format (TIFF) is widely used for Images both

black&white and color. It enables storage of color separated RGB and CYMK digital data.

CompuServe came with the Graphic Image format (GIF) which is widely used ion the Internet World Wide Web today. JPEG (Joint Photographic Experts Group) is a compression algorithm which is used by most file formats to compress data. JPEG under lossy algorithm, provides tremendous data compression ratio for low contrast images.

Some of the common file formats are discussed below in detail:

2.9.2 TIFF file format

TIFF is an acronym for Tag(ged) Image File Format. It is one of the most popular and flexible of the current public domain raster file formats. etc.

TIFF is primarily designed for raster data interchange. It's main strengths are a highly flexible and platform-independent format which is supported by numerous image processing applications. Since it was designed by developers of printers, scanners and monitors, it has a very rich space of information elements for calorimetry calibration, gamut tables, etc. Such information is also very useful for remote sensing and multispectral applications.

- TIFF is capable of describing bilevel, grayscale, palette-color, and full color image data in several color spaces.

- TIFF includes a number of compression schemes that allow developers to choose the best space or time trade off for their applications.

- TIFF is not tied to specific scanners, printers or computer display hardware. It is portable across different operating systems and thus has remained the most popular format for cross platform DTP applications.

- TIFF is carefully designed to be extensible to be able to modify and upgrade as new needs arise. The current version of TIFF as released by Adobe is Rev 6.0.

- TIFF enables special purpose formats to be added and thus has resulted in a wide range of TIFF formats that no one application can read all formats of TIFF files.

2.9.3 Postscript Environment

PostScript is a vendor specific format developed by Adobe Inc. It was developed as an image mapping program. It is not a document language (though because of its close association with paper documents has been classified as document processing language). Paper and computer screen has been the most common publishing media. Postscript and its derivative, PDF does a good job on both this media. Each page is treated as one element of a document. Very little control is provided for the actual content and element in a page. Postscript is the most widely accepted and practiced cross platform document format in the publishing industry. There popularity was due to the preservation of the original format of the document.

In a PostScript environment, an Encapsulated PostScript file (EPS) format is used to describe one single page (or an element). A PostScript file (PS) in itself can contain multiple pages (or documents!). Both EPS and PS are codes which enable a PostScript printer to interpret the document on a paper. PDF was a derivative of PS with the ability to support a soft view on the monitor. In addition, effort has been finally made atleast to distinguish some elements of a single page separating the text in a page from the illustrations. This enables retrieval searches and recreation of document content to an extend.

Greater emphasis is given on PDF documents for image compression. Different algorithms are used on different identifiable elements of a page to reduce the digital file size.

2.9.4 EPS file format

EPS stands for Encapsulated PostScript file. It is the most popular format for both vector and raster image format. It's success is primarily due to the revolution created by PostScript in the DTP world.

EPS is in essence a PostScript file except that it holds information about a specific page (unlike PS's multipage document type). EPS files are formatted and ready to be processed by the output device. A boundary code defines the image area and size. Since PS codes can't be interpreted by device drivers, a preview header is attached. All EPS files are not the same as like PS files. Each application creates its unique prolog and is left to the PS RIPs to decode them. Some of the EPS files can be much larger than a similar TIFF formatted file.

Some DTP applications like Illustrator, Freehand creates Vector based EPS files which are the best choice for an image format as it preserves all the information in a coordinate system for the printer to decode to its specific capability. Applications like Photoshop has to create a Raster Image in EPS format, while applications like Corel resort to a combination of both.

With PS Level 2, the image formats has been extensively enhanced to handle the various color spaces, though EPS is primarily targeted for CYMK and RGB.

EPS file format is also used to store separated CYMK image. A special format called DCS (Desktop Color Separated) is used to carry five channel information as separate images. It includes the four color separation and a composite image. The composite image is used for OPI (Open Prepress Interface) image manipulation and is substituted by the separated images at the time of RIPping.

2.9.5 JPEG file format

JPEG as good as a file format is also worth considering as a compression format. TIFF and PDF (the new wave for PostScript) offer JPEG compression.

As Internet (Digital Publishing) is fast catching up as a lucrative business for publishers, image formats like JPEG and its derivative JFIF are trying to emerge as universal interchangeable file formats.

Media like CD–ROM publishing, image archives, digital video compression are the best candidate for JPEG file formats. Hence, in view of the emerging trend in

publishing, JPEG and GIF file formats are discussed in this paper with equal importance as TIFF and EPS file formats.

JPEG stands for Joint Photographic Experts Group. The compression method invented by this committee is widely used today, most commonly in the form of JFIF (JPEG File Interchange Format) files. The committees goal was to reduce the file size of natural, photographic-like true-color images as much as possible without affecting the quality of the image as experienced by the human sensory engine.

The subject matter best suited for JPEG compression are the types of images found in nature. Sharp, contrasting edges are not common in nature, and they do not fair well when JPEG compression is pushed to its limits. Subjects consisting of black lines on a white background degrade rapidly. Similar things such as typefaces and line art are poor subjects for JPEG compression as well. Water, sky and skin can be generously compressed with the minimum of loss and retain their rich, true colors. We humans perceive the images around us in a reasonably well understood way. Science understands the mechanisms of sight and the psychophysics of visual perception. It is this aspect of our perception that JPEG compression exploits in an effort to reduce the demands on our modems and hard drives.

While most image file formats use an RGB (red, green, blue) value to describe each pixel value, the JPEG format converts this data to luminance (brightness) and chrominance (hue). This allows for separate compression of these two factors. Since the luminance is more important to our senses than the chrominance, the algorithm retains more of the luminance in the compressed file. The JPEG compression algorithm works on individual groups of 8 x 8 pixels. It calculates a Discrete Cosine

Transform (DCT) for the entire group, quantizes the DCT coefficients, and then applies a Variable Length Code compression scheme to the coefficients. It's the quantization step where the loss of color information occurs. And the DCT is the reason why JPEG doesn't do so well on sharp edges. DCT amounts to trying to represent the image as a sum of mathematical curves. That works great on relatively smooth images, not so well on sharp edges.

2.9.6 GIF file format

With the developments in PostScript Level 2 in terms of supporting RGB and Lab color spaces, JPEG and GIF has been accepted as practically usable file formats in the prepress work flow. Though these formats are originally ment for storage and display, the DTP graphic applications are supporting them as DTP merges with DIP (Desktop Internet Publishing)

GIF stands for Graphics Interchange Format, and was developed by the Compuserve for the whole purpose of networked transfer of digital images across platforms.

Some key features which makes it a unique and valuable format for the World Wide Web include compression, transparency, interlacing and storage of multiple images within a single file which allows for a primitive form of animation. GIF offers more features than JPEG (or TIFF and EPS for that matter).

GIF renders colors with a global/source color palette. It is an ideal format for storing illustrations with limited and critical colors and for images with shape edges

line line art. However, DTP page make layout applications (until recently) do not support GIF or JPEG format as they are not ideally suited for PostScript environment and high definition images.

A good comparison of JPEG and GIF images is published in The Seybold Report on Internet Publishing Vol 1, No 11.[4]

2.9.7 SGML

SGML is a data encoding that allows the information in documents to be shared either by other document publishing system or by applications for electronic delivery, configuration management, database management, inventory control etc.

SGML is an international standard and is vendor independent. It is a mark–up language specifically developed for encoding documents. By tagging data with its role and any other useful identifier, SGML allows information to be readily located and re-used. It is possible to selectively extract documents from an archive and republish it a new media or format. A training multi–media kit can be prepared extracting specific details from a technical reference service manual.

It is the only document format which meets most of the ideal document requirements. However, till recently it has been used only for documents for which content is critical while PS has been the de facto for all format rich publications. Because of its restricted use, it has been a professional tool and not as user friendly as PS environment.

SGML data in its raw form contains ASCII codes and bitmapped image containers. Encoding and compression can be efficiently handled.

HTML and Extended Markup Language (XML) are SGML's ported form for World Wide Web.[5] More work is on going on XML and Channel Defintion Format (CDF) standards.[6]

2.10 Copyright Clearance

Intellectual property protection can be grouped into four category: copyrights, trademarks, patents and trade secrets. Each of these legal concepts protects different things; each leaves different things unprotected.

Copyright covers the expression of ideas, but it does not protect the ideas themselves.[7] An idea can be read and restate in different words. A patent can cover ideas if they are useful and original, but only if they can be classified as machines, processes or compositions of materials. A trademark protects the brand identity in the marketplace, but does little to keep competitors from cloning the products under a new brand name. Trade-secret rules protects employers by binding employees with confidentiality clauses in their employment contracts, but is no guarentee if the confidential information is used through unidentifiable track.

The first modern copyright law, the Statute of Anne, was passed in England in 1709; it gave copyright for 14 years. In 1775, the Universities Copyright Act further spelled out the protection that copyright afforded to British citizens.

The first U.S. copyright law was passed in 1790 which protected published materials like Maps and Books. Amendments in 1870 extended copyright protection to paintings, photographs, drawings, statuary and dramatic performances, and in 1909 it added coverage for unpublished works.

All those laws required that copyrights must be registered before any protection was granted. In 1976, the U.S. made a fundamental change in its approach to copyright. The new law said that copyright existed from the moment a work was fixed in a tangible medium of expression. Unpublished works, and works that had been published but not registered, now had the same legal protection as registered works. As on today the TERM period of the copyright protection stands at the lifetime of the creator plus seventy years.

Copyright protects the expression of ideas, not the ideas themselves. Another person could express the same idea in a different format. This helps a healthy competition in presenting an idea to the world. As such, if there's only one way to write something, that expression cannot be copyrighted. By Law, copyright protection materials can be copied in small quantity for academic and evaluation purpose. Copies for personal use is widely accepted.

On the Internet the issue of copyright has created lot of concern. New software products are being developed which provides watermark on the digital files. The Library of Congress is developing a means of registering copyright and related documents over communications networks, including the Internet.

The Copyright Office Electronic Registration, Recordation and Deposit System (CORDS) is currently in the prototype and testing stage. When completed, copyright holders will be able to electronically register copyright for printed and material works, as well as submit copyrighted works in electronic form for inclusion in the Library of Congress.

The "Copyright Clearance Center" is a Internet Web Site which acts as an authority to issue copyright for electronic document repository and handles royalty issues for those who demand an electronic copy of the document in their repository.[8]

2.11 Security Issues

Security issues over Internet is two phased. The security of protected documents and the security of critical information transferred over public Internet network.

With the overwhelming advantages of TCP/IP protocol, Corporate companies are also adopting it for internal Local Area Networks. Within the confines of a single LAN intranet, Internet technology carries no more security concerns than any other networking protocol. However, when a company connects its LAN to the Internet, it must begin to take external security into account.

The major advantage of Internet is its economical means of connecting the global operations of a corporate company. The use of public telecommunication line becomes a security threat.

Firewalls for intranet and data encryption are commonly used to secure the confidential data and information over a public network.

2.11.1 Firewalls

Firewalls are designed to limit the access that outsiders will have to a company's resources, while giving company employees reasonably complete access to the public resources of the Net.

Web Server in its simplicity is a file server adopting the HTTP (Hyper Text Transfer Protocol) standards to communicate with the clients and other servers on the Internet. Files of type text/binary with certain MIME types like HTML, Plain text. GIF images. JPG images and the like can be transferred between a Server and a Client.

The simplest form of firewall is a packet filter in the router that connects your LAN to the Internet. A packet filter examines each IP packet as it arrives at the router and determines whether to pass the packet through or block it. It makes the decision by looking at the source and destination address fields of the packet header, comparing them against a list of rules that the network administrator has written.

For example, an intranet with IP address ranging from 129.21.0.0 to 129.21.255.255 knows that any access from a source in that range is from with in the LAN. And if the requested URL is within the same range, then it serves the page if not the request is rejected.

Packet filters are not completely secure. With enough understanding of the TCP/IP protocol hackers can penetrate the Firewall. It is debated as risk prone as any

other secure documentation. It is a compromise between the load on the server and the amount of security required.

Proxy servers provide even greater security. Additional software can verify that the messages addressed to a server really are using the correct service protocols and are not trying to jam bad data into the machine. This is the job of the proxy server or application-level gateway. A proxy server has to perform almost all of the operations that a real server would, except that it doesn't actually provide the service. If an incoming message passes the proxy's inspection, it will be forwarded to the real server for execution—where the message assembly, syntactic parsing and session state tracking will be done all over again.

Proxying is much more CPU-intensive than packet filtering, but it is less demanding than service fulfillment. A proxy can therefore be set up to track several protocols and protect several servers.

2.11.2 Data Encryption

For many companies, the Internet is an ideal data transportation system—or it would be, if only the data could be kept secret while traveling over public wires. The Net is ubiquitous and cheap, but it is not trustworthy. The obvious solution is to encrypt the data during its journey.

Different encryption schemes are available for commercial use with multiple tradeoffs among cost, speed and security. U.S. law does not allow exporting any

encryption keys longer than 40 bits. Hence there are issues of differences and incompatibilities when data has to be communicated across the globe.

Data can be encrypted by hardware devices or through software.

Hardware devices are attached to the IP packet routers of the companies Internet server. Data going out to the public network is encrypted using special algorithms. A key is passed to the receiving station on the other end of the companies network. Once accepted, the encrypted data is tunneled through the public network. At the recieving end, a similar device decrypts the data based on the key exchanged. The encryption algorithm and/or the key may be constantly changed to keep hackers from decrypting the data hacked at the public network.

When specific data has to be protected, Software can be used who's encryption action can be programmed to suit the level and layer of protection required.

End Notes for Chapter 2

1 Kinsley, Lawrence. *Consumer Behavior in Online Publishing.* The seybold Report on Internet Publishing, Volume 1, Number 5, pp. 9-14. Seybold Publications, 1997.

2 Kinsgley, Lawrence. *Comdex Foretells of Web Printing.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 17-21. Seybold Publications, 1996.

3 Iversen, Osmund. *Server Push or Client Pull.* The seybold Report on Internet Publishing, Volume 1, Number 2, pp. 39-40. Seybold Publications, 1996.

4 Houldsworth, Andy. *Tuning your Imaes for the Web.* The seybold Report on Internet Publishing, Volume 1, Number 11, pp. 12-16. Seybold Publications, 1997.

5 Walter, Mark. *W3C publishes draft of simplified SGML.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 3-5. Seybold Publications, 1996.

6 *Microsoft pushes Channel Definition Format.* The seybold Report on Internet Publishing, Volume 1, Number 8, pp. 25. Seybold Publications, 1997.

7 McKenzie, Matt. *Copyright Protection: Understanding Your Options.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 6-14. Seybold Publications, 1996.

8 Copyright Clearance Center. World Wide Web site http://www.ccc.org/

Chapter Three

On-Demand Digital Printing

Digital Printing refers to the technology of producing images on paper directly from a digital image without going through the conventional film and platemaking (including CTP) process.[1]

On-Demand extends and narrows this capability to system which print a required title (or book) with run lengths as low as ONE copy.

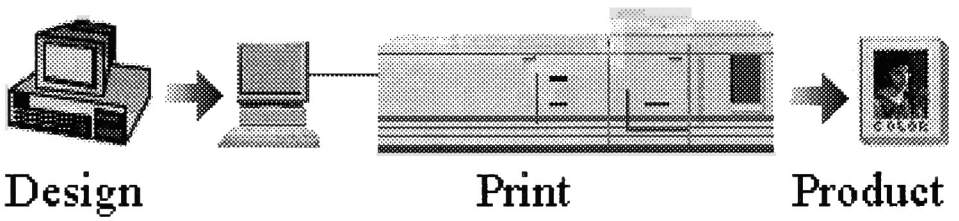


Figure 3.1, Work Flow for Digital Print Production

The concept of On-Demand digital printing, as stated above can literally include all the Digital printing machines available in the market today starting from a photocopier with digitizer (like Canon) to a high–speed conventional printing machine with a digital imaging drum (like Heidelberg GTO-DI).

Another new feature of (digital) printing, which is gaining momentum in publishing domain, is the concept of variable printing. Variable printing encompasses printing with the possibility to change the data on a printed area (like serial number,

names, images, address etc..) on every copy or over a short run length. Printing of envelopes from a database using MicroSoft Mailmerge in a laser printer is an application of variable printing as well as the printing of Lottery tickets using a traditional platen machine. However variable printing in today's discussion is more centered around variable digital printing dealing with digital image manipulation which enables variable text and/or images to be printed on each printed page.

Some examples of the common printing machines and its category listed in the table will give a better understanding:

Machine	System	Category
Heidelberg SM102V	Conventional/Digital Metal Plates	Traditional Printing
Heidelberg AUTOPLATEN	Variable Images on different Polyester Plate	On-Demand Printing
Heidelberg QuickMaster DI	Digital Plate Exposure	High Speed Digital Printing
Canon CLC 800	Work Station with print spooling	On-Demand Digital Printing (Copier)
Xeikon DCP-1	Digital RIP / Image server	On-Demand Variable Digital Printing

Table 3.1, Conventional and Digital Printing Systems

3.1 Necessity for On-Demand Digital Printing

Success of On-Demand digital printing is because of the following three reasons as explained below:

a. Cost effective for short runs

b. Data archiving

c. Easily portable to other publication media

3.1.1 Cost effectiveness

In true business sense, On-Demand digital printing offers major savings in the business of book printing. Following is a comparative study of the cost involved in producing a typical 192 pages black & white book.

Specification

Trim Size	8.5" x 11"
Quantity	100cp
Binding	Soft Cover
Finishing	UV Varnish on Cover

COVER		*TEXT*	
Pages	4pp	Pages	192pp
Paper	80# Cover Coated	Paper	60# Text Smooth
Color	4Col on Front/Back	Color	1Col throughout

The cost of production is estimated on a cost center basis. With digital technology some of the cost centers differ from that of the conventional production. Following table summarizes the estimated cost in US dollars.

The conventional Press unit is assumed to use a 8up B&W perfecting press for this job with a automated perfect binding machine. The digital printing unit is assumed to have a B&W high speed digital printer and a 4–Col digital printer, with in–line softcover binding machine. The system configurations will be discussed later in this chapter.

	Conventional Production	Digital Print Production
Creative Design	1,000	1,000
Artwork Preparation	2,800	2,800
Film Output	1,100	-
Separation & Proofing	120	-
Film Assembly	200	-
Plate Making	520	-
Make Ready & Print	820	1,800
Paper	432	300
Finishing & Binding	400	200
Total	7,392	6,100

Table 3.2, Cost Comparison of Conventional and Digital Print Production

The above prices are estimates worked out by a printing firm which compared its operational cost against the digital technology. The data are for comparison and should not be taken as any measure of actual cost

The cost saving is typically over 20%.

3.1.2 Data Archiving

For a printer, storage of films and plates are true nightmare. Properly organized storage system needs expensive space. Reprints could be done from the artwork archived as application files in electronic disk. However it calls for redoing the film and film layout. As this involves cost and time, Printers in most cases store them as imposed films.

Retrieval of these films is labor intensive and periodic maintenance of the films with proper coding and auditing is required. Being delicate, films need protected storage cabinets, which add to the space and cost.

Digital printing technology requires the originals to be maintained in the digital format (either as application files for easy modification or as RIPed files). Digital files provide great flexibility in archiving and retrieving. In addition, electronic transfer enables files be stored at even remote places. It occupies less space and maintenance time.

In long term, this could be visualized as potential saving in the operational cost of the press.

3.1.3 Portable to other Publishing Media

Publishers are being wrapped around with the novelty of the new media revolution. It will not be long when most of them will start using print as a supportive medium rather than as the primary media of publishing. When this happens, the original data has to be compatible through the different publishing media.

Digital printing provides compatible front end processing to accept most of the common file formats. Being digital, these front ends can be easily upgraded and modified as and when the technology changes. By directly accepting these files, the upgrading of other prepress equipment like imagesetters, platesetters, color separation and proofing systems can be avoided and cost saved.

Alternatively, the files, which are created primarily for digital printing, can be easily ported to other media with very little conversion. This enable the data to be distributed globally and constantly updated with very little material cost. Hence, every reprint of the production can be done with the complete updated information with little or no extra amendment cost on the printer.

3.2 Concept of On-Demand Digital Printing

The concept of On-Demand digital printing is discussed in the following category:

a. Document manipulation

b. Document distribution

c. Printing engine characteristics

d. Finishing Line

3.2.1 Document manipulation

The work–flow for all digital printing machines start with the origin of digital image. The most common format and mode in which the images reach a book printing house from the artwork/design department will be:

File formats

a. Application file like PageMaker

b. Postscript multi–page document

c. A PDF(Acrobat) file

d. Digital printer specific file like DocuTech Prn file

Transfer Mode

a. Transfer Media like MOD, Syquest, ZipDrive

b. Modem transfer through Standard or ISDN Line

c. Attached file through Email

d. FTP transfers

In contrast to conventional printing, digital printing enables a direct link to the client through a File/Print server. Once a standard format and transfer mode is established between a publisher/author and a printer, the ready–to–print files can be downloaded to the file server of the digital printing machine directly. By calling in a standard publication setting (size, paper, pagination, blank pages etc.), the operator can divert the file for printing.

This, in practice, is an explicit example of the high productivity and time–to–market achievable through digital printing.

Digital printing front–end can be visualized as an extension of the Print Dialog Box of standard desktop publishing software. Additionally, electronic pagination and simple printer setting commands completes the front–end operations.

Looking at the technology point of view, the front–end of most of the digital printing machines are very powerful, multitasking file servers and workstations considering the huge amount of digital images it has to crunch and pass to the printing engine. The discussion of the mechanics of the front–end-system is beyond the scope of this paper.

3.2.2 Document Distribution

Again, digital documents bring a school of thought in the way documents are handled. In conventional printing, the documents (may be the application files) are sent to the printer/repro house. They are processed in the prepress and printed as per the forecasted quantity. The printed books are stored and distributed to various places as required.

With a digital printing press, 'print–and–distribute' has been revolutionized to 'distribute–and–print'. The digital documents can be stored in a central Image server. As and when the need arises these documents can be down loaded to a digital printing press and exact amount of copies printed. The documents in this way can be distributed even across the globe to a place where the printed copies are requested. In essence printing is done on demand rather than as forecasted. Internet and Bulletin Board services provide excellent resource for distribute and print systems.

The printer can concentrate on his core business without being compelled about storage, obsolescence stocks and freight forwarding.

A word of caution here, with global access, the publisher is in a position to choose different printing firm in different geographic areas to undertake his job. With no extra cost incurred in artwork/film/color proof/plate, publishers can choose different printers for their reprints. With proper business ethics, these concerns/threats should be mutually discussed in the business contracts initially.

3.2.3 Print Engine characteristics

The core of the digital printing concept is in the technology behind the printing engines (machines). The most common principles used are electrostatic and ink–jet imaging. Machines come in varying formats and capabilities and are suited are specific target market. The technology is still undergoing extensive research and new improvements are introduced almost every year.

Substantial materials are available which discuss in detail the working principle, pros & cons of the various digital printing engines available in the market today. A good reference source will be the book "*On-Demand Printing - The Revolution in Digital and Customized Printing*", written by Howard M. Fenton and Frank J. Romano and Published by Graphic Arts Technical Foundation.[2]

The same book is also recommended for an understanding of the principle behind the color digital printing machines. This report only elaborates on the monochrome digital printing options pertaining to book publishing industry.

The following is a summary of the features of black & white digital printing machines from Major players in the market.

Product	Manufacturer	Print Techn.	Print Resol.	Paper size
DocuTech 4635	Xerox	Electrostatic, Hn Laser	600dpi	14"x17"
Lionheart 6092 LX	Kodak	Electrostatic, LED	600dpi	11"x17"
IBM	IBM	Electrostatic, LED	600dpi	12.6" web
Imagefast 850 IE	Dephax	Ion-Beam Tech.	300dpi	8.5"x11"

Product	Speed	Duplex	Postscript files
DocuTech 4635	135pg Ltr	Using same unit	Fiber Gateway (optional)
Lionheart 6092 LX	92pg Ltr	Using same unit	Postscript Printer
IBM	300pg Ltr	Simultaneous	Print Services Facility
Imagefast 850 IE	425pg Ltr	Simultaneous	PS RIP (Vista Server)

Product	In-line Binding	Apprx. Cost
DocuTech 4635	Saddle & Perfect Bind attachment	$418,700
Lionheart 6092 LX	Saddle Bindattachment	$194,000
IBM	Third Party solutions	~$675,000
Imagefast 850 IE	Third Party solutions	$1,500,000

Table 3.3, Comparison of Digital Printing Systems

3.2.4 Finishing System

Surprisingly until recently not many developments were done in the Binding machinery to go along with the speed of the On-Demand digital printers.

In contract to conventional printing, digital printers can print one copy of a book beginning with its first page to the very last page, before proceeding with the next copy of the book. Unlike a 16 pages signature, almost all digital printers print a max of four letter size pages.

This pose new challenges to finishing machines. A machine which can go in-line with the digital printer, collate the sheets as it comes out, wrap the cover (mostly feed separately) and bind them either as saddle stitch or perfect bind and finally trim

them.[3] The conventional high-speed bindery machines will be of no use for such applications.

Xerox, supports in-line saddle stitch binder, in all its DocuTech printers. There are quite a few third party suppliers of in-line and off–line perfect gluer, saddle stitcher, thermal cover laminators and spine printer which needs careful evaluation before purchase commitment.

Perfect Binding

The most common print-shop oriented approach is softcover perfect binding, involving a wraparound cover. Perfect binding in the print shop is far faster than necessary for on-demand printing. Even top-of-the-line DocuTech, at 135 pages per minute, takes atleast a couple of minutes to print an average book of 100 pages. The print-shop perfect binders and three knife trimmers are much faster for this.

Another consideration with print-shop binding systems is its setup time, complexity and wastage. Even the most automated machines will waste a few copies during setup. With On-Demand printing of one or two copies these wastage might amount to 100-200%.

Xerox, with collaboration of third party manufacturers, have a few solution for this requirement.

* In-line perfect binder for DocuTech from Bourg
* Offline perfect binder from Planax
* Offine adhesive binder from Powis Parker

Three Knife Trimmer

Books finished with perfect binder need to be trimmed to final size. In conventional signature binding, trimming is a must to separate the pages. In digital printing, trimming is done to even out the pages. It should be remembered that almost always the printing is done on exact page sizes (letter size) and trimming makes the book smaller than the actual size. When size is critical, bigger pages need to be used with allowance for final trimming

The Challenge Docutrim is a low-cost low-speed alternative to a three knife trimmer. The machine uses one knife and rotates the book two times to cut all its three sides. Challenge Model 20 is a low end machine with manual settings.

Saddle Stitch Booklets

For publications which are more for internal use and are much less in page extent, the best and cost effective form of binding will be saddle stitch binding.

Xerox DocuTech comes with standard booklet maker. The pages are collated folded and sent to the stapling unit. The preprinted cover (optional) is feed from a cover feeder and is stapled together with the rest of the pages. An on-line trimming unit, trims it to exact size set.

Being in-line with the printer and with easy setup functions, it is the fastest and easiest form of binding for on–demand publications.

Hard Case Book Binding

When a publisher wants to have a few copies of his hard bound publication, next to hand bound, the best way is to use Book Station supplied by Flesher Corp. For printers who service publishers, this machine will be very useful. It fills a dream niche of a publisher, who always like to have a few copies at no risk to send as sample copies for market demand forecast.

Finishing Machines

For commercial printers, it is very common to have the covers laminated. With on-demand printing, the cover is mostly color photocopied or printed on color digital printers. To go one setup further to have a in-house laminator GBC sells a single/double side Impression Maker which uses thermoset lamination foils to seal the cover.

Spine Printer

Most of the perfect bound books carry text in their spine. Xerox offers a Spine Printer, which prints on the spine of a finished book using Ink-Jet printer driven from a DOS/MAC PC. Hence variable printing on the spine with selective names, serial numbers are also possible.

Endnotes for Chapter 3

1 McCalpin, William. *Why Every Electronic Print Architecture is Wrong.* The Xplorer, Fall 93, p. 4. Xplor International, 1993

2 Fenton, Howard M. and Romano, Frank J. *On-demand printing: the revolution in digital and customized printing.* Graphic arts technical Foundation, 1995.

3 *On-Demand Binding of Books.* Seybold Report on Publishing Systems, Volume 24, Nuber 4. Seybold Publications, 1996.

Chapter Four

On-Demand Digital Publishing

4.1 Definition

Publishing is a business concerning information dissemination. All these years, printed media has been the major media for publishing.

On-Demand Publishing is a relatively new concept referring to the possibility of meeting the customers' need in:

1. providing the selective information requested
2. exact number of copies ordered (any where from 1-100 copies)
3. at the requested place and time
4. at reasonable price compared to conventional book

Study on on-demand publishing is grouped under the following headings:

1. Long awaited publishing solution
2. Revolution in On-Demand Publishing
3. System configuration

4.2 Long awaited Publishing solution:

It has become customary to put the strain on the customer (the reader) to hunt for the information he needs among the huge volume of publications in the bookshops.

It is often true that a reader picks up a book on recommendations, author's popularity or publisher's reputation. When comes to books of academic interest, the above filtering may not lead to an ideal find.

On the other hand, the publishers were short handed, as the most versatile media they use, namely printed media, had no alternative solution to this problem, to an extend that most Publishers and even consumers were unaware of the strain in searching for proper material.

Added to this, the publishers had their own business hurdles.

- When they compile and edit materials for a book, they have to compromise between the amount/depth of information against the consumer segment.

- For most Publications, decision on first print run is always on a crude forecast.

- When they need to produce books for market survey or for presentation copies, they had to make a minimum run for economic reasons

- Academic books get easily outdated and need constant revisions. This keeps the run lengths smaller even if the market is big.

- When demand exceeds forecast marginally, publishers are faced with the dilemma to reprint an economical quantity or to reject the request for the book.

These are some of the concerns, which successful Publishers have learnt to live with. In a situation where the service provider (publisher) and the end user (reader)

are having constraints, they tend to compromise on the content and cost of the published material, which is an unhealthy market position.

On-Demand publishing pushes its way to resolve all these constrains.[1]

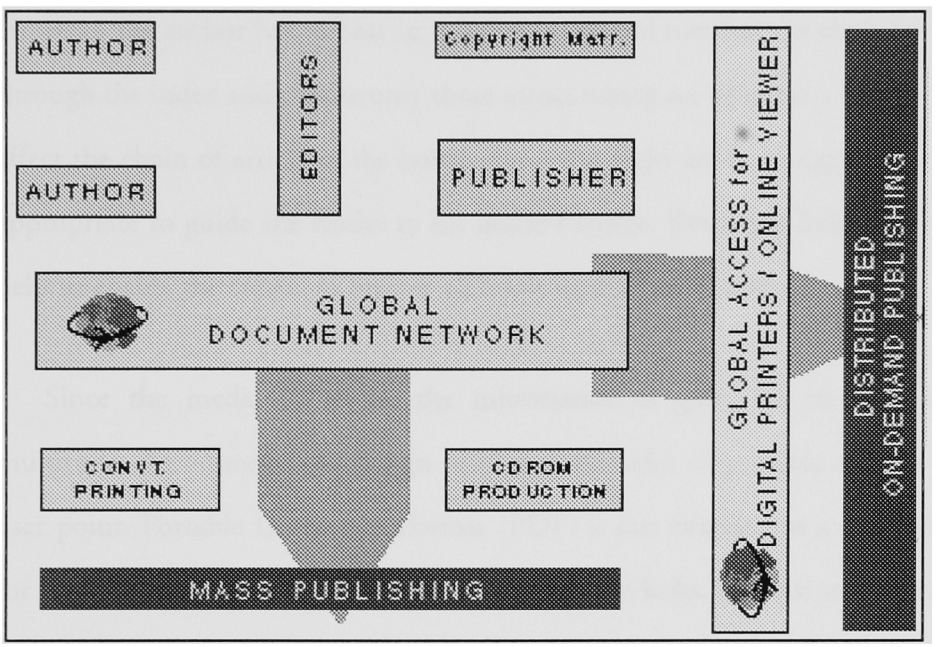


Figure 4.1, Flow Diagram of On-demand Publishing

4.3 Revolution in On-Demand Publishing

On-demand publishing primarily uses on-demand digital printing and on-line viewing technologies.[2] There are three basic modules of this publishing system:

1. Content creation
2. Information dissemination
3. Copyright/Commerce

Original materials are created in much the way as conventional publishing. However the author has to bear in mind that the end user has the choice of browsing through the index and choose only those topics which are of interest to him. This may affect the chain of article in the book. Hypertext links are to be carefully used where appropriate to guide the reader to his desired article. Reference links can be of great help to enable the reader to browse all other related publications.

Since the media in which the information is going to be disseminated is multitude, the editorial and design work mus consider all possible options from end user point. Portable Document Format (PDF) is fast evolving as a common ground for viewing and printing of documents. They enable links, indexed search, sorting and collating of a huge docubase. Publications created in such format can be archived and distributed for print or viewing only on demand.

The common fear for content providers is the issue of copyright as in the vast computer network there is potential chance for information being downloaded illegally. The author/publisher may loose their revenue in such cases and there is the danger of illegal modification of the original content. Almost all countries' federal law makes downloading unauthorized information, as criminal offence.

One way to reduce this threat is by having an association to authorize authors, publishers, distributors, printers, institution, organizations interested in providing &

retrieving information and independent members. This association on a global level can maintain:

1. all original contents with strict confidentiality
2. index articles and make it full text searchable
3. process retrieval request from bookseller, distributor or individuals
4. download information as requested
5. monitor the financial transaction between the end user and the content provider
6. protect the interest of the registered members by constant monitoring of the network

The Copyright Clearance Center makes one such attempt. More global work is required to create awareness to enable distributed demand publishing.

4.4 System Configuration

The configuration suggested here is ideal for a b&w publications with color cover.

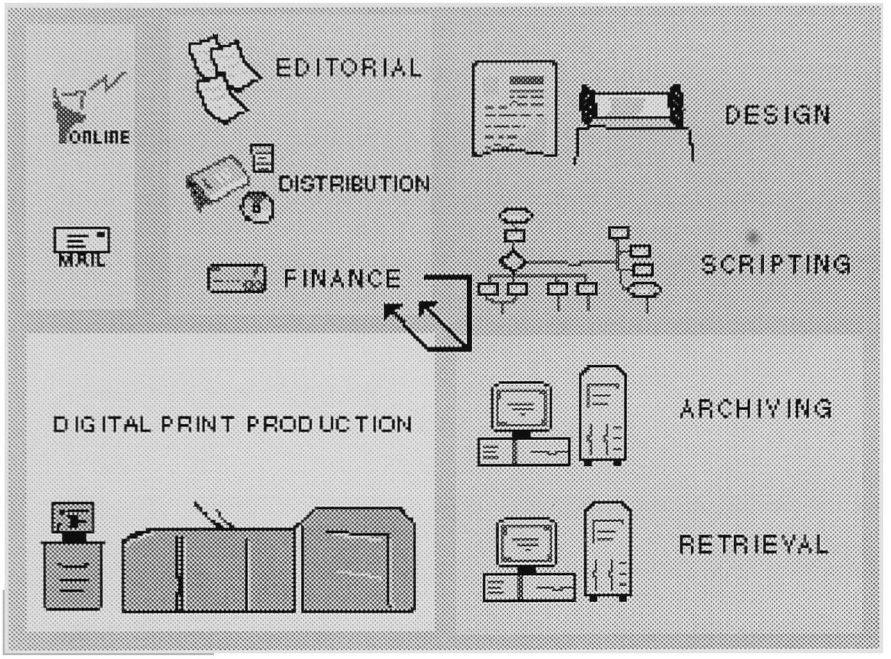


Figure 4.2, Flow Diagram of On-Demand Publishing System

A typical example will be book on prehistoric excavation.

- Content can be varied including entertainment information, scientific information, genetic research details, entertainment context and so on. With on-demand publishing in mind, the publisher groups them into various broad segment (all for one of his title), an impossible task for conventional publication. The material can be collected from various authors as diverse as the category.

- Editorial work has to be done with great care in overlapping. The authors interest must be protected as much so the flavor of each category. Being the same topic lot of resources like pictures and data will overlap.

- Design work is very crucial. The complete document is assembled as raw text with bookmarks. They are indexed and identified under suitable classification. A boiler plate (or template) is designed which can hold the information extracted at later stage by the user can fit in. The template can be common for printed and on-line viewing. Essentially, design in such situation goes in hand with computer programming or scripting.

- The raw content, resources and templates are archived in a digital storage media/server. The major difference from conventional publishing can be seen here. The raw content is separated which allows for constant update, selective publishing and searching. This feature, with a little extra effort, has great potential for continued publishing and will open up new market.

- Document distribution can now be performed in a variety of ways. Upon request for a entertainment version of the title, the search engine can gather selected information from the common database of the raw content. This content can be placed in the appropriate templates and sent to the requestor. Another attractive approach will be to provide on-line interactive client software through the World-Wide-Web. This enables the user to make his own recipe. The customer can choose an amateur version with entertainment content for specific topics, make his own presentation or use existing templates and download the files (all such transaction can be governed by an association as discussed above).

- If the market is not matured enough to make use of such technology, we are not lost either. The raw content and templates can be grouped for commercial printing as well. Digital printers are easily networked globally. Hence, the publisher can have a bulk printing of a reasonable quantity and offer remote digital printing globally, by linking up with existing digital printers.

Endnotes for Chapter 4

1 Weinberger, Dr. David. One-to-One Publishing. The Xplor, Fall 95, p.8. Xplor International, 1995.

2 Romano, Frank J. *Digital media: publishing technologies for the 21st century.* Micro Pub. Press, 1996.

Chapter Five

On-Line Document Management

The two core areas of On–line document management are distribute–and–print technology and document storage and retrieval systems.

5.1 Distribute and Print Technology

Telecommunication, file servers and demand printing technologies are growing so rapidly to realise the concept of distribute and print.

This chapter looks in-depth the solution offered by Xerox (who is the leader in distribute and print technology) and some notes about other systems which are fast catching up.[1]

Xerox offers five different services in this area.[2]

5.1.1 Documents Direct

This is an outsourcing service designed to reduce shipping costs for companies that deliver large volumes of documents to multiple sites. With this service, businesses deliver documents electronically to multiple sites. Documents are printed and bound at the remote sites using DocuTech systems and companion finishing equipment. Xerox takes care of setting up and maintaining the network and the printing system.

5.1.2 Xerox Distributed Print on Demand (XDPOD)

Document Direct is aimed at businesses which have large volume of transactions to justify paying to Xerox for maintaining its transactions. The next step is a transaction-based service called Xerox Distributed Print On Demand (XDPOD), being developed in conjunction with AT&T.

This service, designed for DocuTech users, enables them to send DocuTech print jobs to other DocuTech sites around the globe through the AT&T network. The job can originate in any application that can be handled by the DocuTech RIP. DocuTech Job tickets are automatically fed into the XDPOD system. XDPOD system can send jobs from one point to multiple sites simultaneously.

The service provides secure communication, guaranteed delivery, job tracking and notification of job status. While AT&T handles the communication systems and Xerox the server and software, the customer just have to maintain the XDPOD user module.

Compared with sending Syquest cartridges or Magneto Optical Disks to remote printing sites, XDPOD could still be less expensive and faster.

5.1.3 InterDoc

XDPOD lets you pay as you go, but there is still the fee for using AT&T high speed network. It also requires Xerox specific job tickets. InterDoc opens up this format to the World Wide Web users. It enables inplant print departments, commercial printers

and quick printers to set up shop on the Internet and to receive and print digital files electronically from their customers.

InterDoc has two components.

At the Print shop, it provides a complete Web Server that has built-in software for communicating with customer and managing customer-submitted files. The Web home page lists the services the printer provides, makes pricing and cost estimates available, and provides forms for cutomers to identify themselves to the printer and enter job orders.

The customer, using a forms-capable browser, fills out the job ticket form and sends print-ready documents by electronic mail to the service provider.

For DocuTech users, InterDoc provides the universal job ticket to handle networked Postscript files. Xerox provides access to Oracle database from within InterDoc to enable job tracking and report generation. Xerox also has a Preflight validator to check PostScript files before sending to the printer.

5.1.4 Document Access Remote Transfer (DART)

A less elegant, but less expensive method of accepting remote job requests is Document Access Remote Transfer (DART).

DART package has two components. The receiver software runs on Macintosh computer and supports modems upto 28.8kpbs. The sender software is a freeware program and runs on Macintosh or Windows PC.

DART uses MicroBeam's ASAP document transfer program that provided lossless compression for transmitting documents on a dial-up basis. It provides simple drag and drop interface. The program initiates compression, makes phone call, automatically sets appropriate communication protocols to confirm with the receiver and transmits the document to the print shop.

5.1.5 DocuWeb

DocuWeb is targeted at education centers where it can be used to make reserved readings and out-of-print books available to students in print form. The same concept over World Wide Web can be used for other similar applications as well.

DocuWeb is an extension of XDOD system. Users with XDOD workstation search and retrieve documents, view them on the XDOD terminal and initiate print requests. With Web server plug-ins, the users an use any web server as the querying tool.

Xerox Distributed Print-on-Demand Product Comparison:

Category	Type of offering
Documents Direct	Outscourcing either the entire publishing and printing or a specific process to Xerox
Xerox DPOD	Pay-as-you-go access to a network service for remote printing
InterDoc	Web server for remote job submission
DART	ASAP software by MicroBeam for dial-in remote job submission
DocuWeb	Upgrade to XDOD library for Retrieval and initiating printing

Category	Client
Documents Direct	Companies with volume who can commit on outsourcing
Xerox DPOD	DocuTech Customers only
InterDoc	DocuTech Customers (through WWW)
DART	DocuTech Customers (through dial-up)
DocuWeb	Customers managing large Docubase

Category	Network system
Documents Direct	Xerox network utilizing AT&T, Sprint and MCI
Xerox DPOD	AT&T store and forward network
InterDoc	Voice line dial-up or dedicated internet line
DART	Voice line dial-up
DocuWeb	Voice line dial-up or dedicated internet line

Table 5.1, Xerox Print-on-Demand product comparison

Another well known player in the area of distribute-and-print is Adobe. As Adobe made a revolution with the introduction of Postscript making it a universal print platform for graphics and text, it is bring out Acrobat as an equivalent in the Electronic Document format. Acrobat, interestingly provides a common platform for printing as well as viewing on computer screens be it in Web browser or exclusive software. It also makes document platform independent. Hence, a Acrobat file created from a Macintosh application can be viewed on a Unix platform and printed from there on any postscript printer over a network.

Acrobat does it with its Viewer which is a freeware available for Macintosh, Windows and Unix platform. With its Acrobat Exchange product, Adobe provides an excellent solution for document management, retreival, search and print. Being page oriented, it is also possible to make up ones own selective article from an archive and print it on-demand.

5.2 Document Storage and Retrieval systems

Before discussing further some clarification is required on the various retrieval models being used in today's electronic information era.

- DataBase of fields like employee details, share market data
- Image Archival like Photo Library
- Full Text based search functions like Research papers
- DocuBase which are primarily page oriented retrieval systems

For Print-on-Demand and Distribute-and-Print system, we are interested in the products with DocuBase management based on full text search facility. Of a few software available in the market Acrobat from Adobe is the leader with the primary target of digital publishing.

Adobe Acrobat Pro is a complete software package that comes with the following programs.[3]

5.2.1 Acrobat PDF Writer

A driver that enables you to "print," or create, PDF files from common business applications, such as word processing and spreadsheet programs. This invites the whole world of Document creators to get into the Acrobat wagon as easily as they did during Postscript revolution, ensuring complete portability.

However, for critical conversion, it is best to create a PS file and convert it to PDF using Acrobat Distiller which is also bundled with Acrobat Pro.

5.2.2 Acrobat Search

Provides full-text search capabilities for PDF files that have been indexed with Acrobat Catalog software. A feature which enables document retrieval and printing much practical even over a huge data archive.

Adobe Type Manager

ATM - A utility that gives you sharp, clear text on-screen and in print, at any size and resolution.

5.2.3 Acrobat Reader

Enables Macintosh, Windows, DOS, and UNIX users to view, navigate, and print any PDF files they receive.

5.2.4 Acrobat Exchange

Lets you add value to electronic documents and share them with other Acrobat users. You can not only view and print but also annotate, build navigational links into, and add security controls to PDF files. This is an useful tool especially for on-screen readers where the user can be conveniently navigated through an article using the Article link feature.

Acrobat **Distiller**

Helps to convert any PostScript language file into PDF. Use Acrobat Distiller with files from drawing, page layout, or image editing programs; documents containing high-resolution or Encapsulated PostScript (EPS) language artwork or images; or documents containing complex blends or gradient fills.

5.2.5 Acrobat Catalog

This software lets Macintosh and Windows system users create full-text indexes for collections of PDF files shared over a network by Macintosh, Windows, and UNIX users, who can then find information instantly with Acrobat Search.

For more details browse http://www.adobe.com/acrobat/prodinfo.html.

Endnotes for Chapter 5

1 *Orlando Exhibit Review.* The Xplorer, January 1997 Edition. Xplor International, 1997. World Wide Web site http://www.xplor.org/

2 *Distribute, Then Print: Global Networks Take Demand Printing to remote Site.* Seybold Report on Publishing Systems, Volume 24, Number 22. Seybold Publications, 1996.

3 Ames, Patrick. *Beyond paper: The offical guide to Adobe Acrobat.* Adobe Press, 1993.

Chapter Six

Hypothesis

The goal of this thesis work is to demonstrate a working business model of remote on-demand printing.

A software product will be developed which will enable printing of thesis and dissertation work from the archives of RIT's Wallace Memorial Library on a DocuTech digital printing device at the Digital Publishing Center at RIT's School of Printing and Management Science.

The hypothesis is to show that the author/publisher can use the Internet and digital printing technology to create a new business model which is almost a necessity for todays networked business world.

The goal will be achieved by the following steps:

1) Understand the Internet architecture including, clinet-server model, HTTP protocol, HTTPD Web server, Common Gateway Interface (CGI) model for Web server and File Transfer Protocols

2) Study various document processing systems including, document creation, archival and retreival. Understand the limitations of file formats supported by Web Browsers and compression aspects for file transfer

3) Learn CGI programming using Perl, C++, Java and Oracle based Web script

4) Update on the trends in the digital printing and publishing world, digtal network of libraries, on-line publishing and Internet commerce

5) Detail working know-how of Xerox DocuTech system and Wallace Library cataloging system

6) Develop a model for remote file transfer to the DocuTech system

7) Develop a business model for on-demand print request

8) Develop a software package which will demonstrate the concept by implementing a practical model which will enable internet users from round the world, order printed copy of a Thesis from the electronic archives of Wallace LIbrary

Hypothesis

An Internet based service can be constructed that will provide anyone with World Wide Web access to easily search for, view an abstract, download full test or place an order for an on-demand printed copy of a graduate thesis, and provide and automated transaction processing system for managing the business, incorporating copyright issues, royalty payment for the author, supporting wide range of digital file formats including HTML, PDF and PostScript and provide facility to support world readable abstract with password protected full content.

Chapter Seven

Methodology

The goal of establishing a business link between Wallace Library and Digital Publishing Center at School of Printing Management and Science, Rochester Institute of Technology, will be achieved in the following stages:

7.1 Converting Thesis

Convert a sample thesis to digital format for archiving in Wallace library's database.

1) Ronald James Recene's graduate thesis work is chosen for this test.
2) A digital copy of his thesis work will be requested from him.
3) A suitable digital archival format will be chosen after discussing with Wallace library Staff. The thesis work will be converted to suitable digital format.
4) A suitable (Web) Server will be chosen to archive the digital file.
5) The naming convention and the directory structure will also be decided after discussing with the Wallace library staff.

7.2 Creating Catalogue Tags

Create Special tags in Library Catalogue to hold the thesis abstract, URL for the digital copy and URL for the print request.

1) The tagging convention of the Library Catalogue system will be studied.

2) Standard tag for large text container will be used for thesis abstract.

3) Standard tag for URL's will be used to point to the digital copy of the thesis.

4) Similar URL tag will be used as an hyperlink to submit necessary details to the DPC server for processing the print request.

7.3 Update Library Record

Edit the record in Library Catalogue for the above thesis with appropriate details in the new tags.

1) The library cataloging staffs' help will be requested to update the record for this thesis.

7.4 Demonstrate File Transfer

Demonstrate File Transfer Protocol to transfer the thesis from Wallace library server to the DPC server

1) Write a program in the Interdoc Server to reflect the URL and title of the item requested by the customer.

2) Extend the program to retrieve the URL automatically and store it in the Interdoc Server.

3) Demonstrate security and authentication restriction in transferring files.

7.5 Interdoc Web Scripting

Demonstrate ability to understand the working of the Interdoc Web Server

1) Write an Interdoc web script to create new job log entry

2) Test to ensure that the script is able to create a genuine job entry with all necessary information for processing a job.

7.6 Interdoc Job Entry

Demonstrate ability to capture specific information for processing a Job using Interdoc system.

1) Study the various database fields of the Interdoc System.

2) Write necessary scripts to capture transaction information from customer including, payment information, delivery address and delivery method.

7.7 Testing

Log the print request in the Interdoc System with pointer to the URL of the actual file.

1) Combine the programs and script from stage 7.4 through 7.6 as a working model.

2) Test if the print request from the Library Catalogue is able to make a job entry in the Interdoc system.

7.8 Business Model

Suggest a business model to handle the financial transaction and delivery handling.

1) Recommend how financial transaction will take place between the customer, publisher, printer and the business center (which can be one and the same).

2) Recommend how the final product will be delivered to the customer.

3) Recommend who in the line will be responsible for customer relations.

7.9 Demonstration and Feedback

Demonstrate the complete business transaction involving the Reader (Customer), Library (Publisher), DPC (Print center) and Book store (Business Center).

1) Involving a volunteer, go through the complete business transaction of requesting a printed copy of Recene's thesis.

2) Document the customer's feedback on the convenience, speed, security risks.

3) Document publisher's feedback on other possible use of this model, copyright issues and security criteria.

4) Document printer's feedback on the software package, job tracking, automated file transfer feature and other potential applications.

7.10 Suggestion for further study

Suggest further developmental works to improve on this model based on the feedback.

Chapter Eight

Result

8.1 Overview of print order processing

This section takes through the steps involved in ordering a printed copy. The user, publisher and printer are on the world wide web potentially in any part of the world.

Author:	*James Recene*
Title:	*Thesis work - 'An investigation into the use of the world wide web as an interface for distributing electronic documents to and from a remote digital color printing site'*
Publisher:	*Wallace memorial library (WML)*
Printer:	*Digital Publishing Center (DPC), at SPMS, RIT*
User:	*Assumed to be a researcher on the world wide web*
System:	*The program developed by the author of this thesis*

Process

1) The user through www search engine or WML search engine, gets to Recene's thesis catalog. User clicks on the "Request printed copy of the thesis" hyperlink.

2) System prompts authentication from user for DPC server.

3) System prompts user for payment and delivery information.

4) System logs the data into DPC database and returns job tracking number to user.

5) DPC staff refers to job entry, opens the url in a browser (or using the geturl.cgi program developed by the author of this thesis) and prints to the DocuTech printer over local network.

6) DPC staff makes financial transaction and delivers the printed copy to the user.

8.2 Conversion of the thesis to PDF

The quarkxpress file from the author (James Recene) was converted to PDF file format. The pictures were sampled at 300dpi and all fonts were embedded.

Section 2.9 discusses the various digital file formats.

Any file format can be used as the end process is going to be a HTTP based file transfer which can support any file type. However, as it is advised to use the www browser at DPC to print the document, it is advisable to use a common MIME type file like text, html or PDF. If the browser at DPC is configured to invoke other common desktop applications like QuarkXpress, PageMaker, GostView then file formats like qxd, pm and ps can also be easily supported.

8.3 Creating hyperlink to DPC

A publisher on the world wide web can use one of the two following formats to redirect his visitor to the DPC system.

Unregistered publisher

By creating a hyperlink with a "GET" http format any publisher (or author) on the net can send details of the url to be printed to the DPC system.

eg:

```
<A HREF="http://129.21.203.108/cgi-bin/dpcetd.cgi?
title=An+investigation+into+the+use+of+the+World+Wide&url=http%3A
%2F%2Fdesign.rit.edu%2Fetd%2Fthesis.pdf">
Request printed copy of this thesis
</A>
```

Note: The hexadecimal conversion of non-alphanumeric characters inside the HREF tag is necessary to parse http request properly.

This is a straight forward approach and needs no special setup at the publisher and DPC server. This approach makes the call transparent and opens the source of the document to an intelligent web user. Hence, this is useful when no special copyright security and royalty management is required.

Registered publisher

To handle the copyright issue, the easiest approach on the net is to protect the document using http authentication method - either by IP address protection or by username/password.

A protected document can be accessed only by authorized user. This enables the publisher to post an abstract of a title for the world and secure the full text document.

When such a protected document need to be printed, the publisher opens an account with the DPC. Details of the document is entered in a database. The system can create a title identification key number which is issued to the publisher.

In case of IP protected documents, the publisher must open their server to allow access for DPC server's IP number.

The print request in this case is created with a "GET" hyperlink carrying the identification number.

eg:

```
<A HREF="http://129.21.203.108/cgi-bin/dpcetd.cgi?rec=Az12T_10">
Request printed copy of this thesis
</A>
```

The DPC system upon receiving the identification number, checks to ensure that the request came from the designated server (this prevents misuse of the identification number), looks up the database for url, title, authentication and logs a new job entry in the system.

The current program doesn't include this module.

8.4 WML catalog record

WML catalog is supported by Innovative Information Incorporation. Following their catalog schema, two user defined tags (856) were inserted into the catalog which holds

the url for the full text version for viewing and another holding the redirect call to DPC.

```
eg:

|uhttp://129.21.203.108/cgi-bin/dpcetd.cgi?|zRequest printed copy
```

The tags inserted will appear when the search returns the catalog record. On a www browser, the tag is hyperlinked enabling the user to choose the redirection. As an experiment, only James Recene's thesis record was altered to include these tags. The hyperlink is of the "GET" call with title and url.

In addition to the 856 tag, a 520 tag was used to incorporate an abstract of the thesis. This provides an overview of the thesis. The full text view hyperlink enables the user to view / download the entire pdf document of the thesis on to a personal computer. As the author of the thesis has not requested copyright protection, no security protection was required.

Marc code of the catalog is available in Annex-I

8.5 File transfer module and security authentication

This program module demonstrates the ability to retrieve a file from the specified url (with authentication if required) and save it in a specified directory on the DPC server.

```
usage:
%> perl geturl.cgi <url> <dir_path> [<username> <password>]
```

The program will attempt to open a http connection to the specified url. If specified, it uses the username and password to authenticate to the remote server. If the connection is not successful, it prints to STDOUT an error with the http error code and error message. If successful, it retrieves the content of the url and saves as a file in the directory specified. The file name is the same as the remote file name.

Runtime output of geturl.cgi is in Annex-II

Code listing of geturl.cgi is in Annex-IV

8.6 Web server

The Interdoc web server is essentially a NCSA httpd server running on Sun Solaris 5.0. The server with the name q662q01.rit.edu <129.21.203.108> listens on port 80.

The document root is /opt/httpd/htdocs/

The scripting root is /cgi-bin/ => /opt/httpd/cgi-bin/

Xerox has added the following alias to print to the interdoc system:

ScriptAlias / /opt/interdoc/public/interdoc/home.html

This changes the default document root and makes the above the default home page for the server. Note the use of ScriptAlias instead of Alias, as the page (like most other interdoc pages) is a script embedded into html tags. This will be dealt in section 8.8.

ScriptAlias /pub /opt/interdoc/public/interdoc

This is the public directory where files which doesn't need authentication are placed. "interdoc" is the program which is called whenever a page from this directory is served.

ScriptAlias /mo /opt/interdoc/private/interdoc

The directory to server protected files. The user can authenticate with their established client account id or use the guest/guest id.

The server is configured to run as username => 'nobody' and group => 'nobody'. Hence all the scripts and cgi programs has to be given read and execute permission for user 'nobody' and read permission for html static pages.

8.7 Project space and files

For this thesis project, to avoid any change to the system / server configuration files, the existing configuration was thoroughly studied and the following directory spaces are used.

cgi scripts /opt/httpd/cgi-bin/

The default NCSA script directory

html pages /opt/httpd/htdocs/etd/

& templates *The default path for NCSA script with a sub directory 'etd'*

interdoc /opt/interdoc/private/en/

The default directory for protected interdoc scripts

Following files are exclusively created for this project

/opt/httpd/cgi-bin/

permission	*user*	*size*	*date*	*name*	
-rwxr-xr-x	mkd8879	1836	Dec 23 21:42	getur1.cgi	- **Annex-III**
-rwxr-xr-x	nobody	4737	Dec 25 17:55	dpcetd.cgi	- **Annex IV**

/opt/httpd/htdocs/etd/

permission	*user*	*size*	*date*	*name*	
-rw-r--r--	nobody	483	Nov 24 18:57	contact.html	- **Annex-V**
-rw-r--r--	nobody	114	Nov 24 18:54	exit.html	- **Annex-VI**
-rw-r--r--	nobody	251	Nov 24 18:56	help.html	- **Annex-VII**
-rw-r--r--	nobody	467	Sep 19 14:37	index.html	- **Annex-VIII**
-rw-r--r--	nobody	243	Mar 26 19:97	mast.html	- **Annex-IX**
-rw-r--r--	nobody	600	Nov 20 13:30	order.html	- **Annex-X**
-rw-r--r--	nobody	696	Nov 24 18:52	selection.html	- **Annex-XI**
-rw-r--r--	nobody	600	Nov 24 18:49	template_index.html	- **Annex-XII**
-rw-r--r--	nobody	854	Nov 24 18:51	template_selection.html	- **Annex-XIII**
-rw-r--r--	nobody	691	Nov 24 18:44	welcome.html	- **Annex-XIV**

/opt/httpd/htdocs/etd/pix/

permission	*user*	*size*	*date*	*name*
-rw-r--r--	nobody	3568	Mar 26 1997	bkg.jpg
-rw-r--r--	nobody	596	Mar 26 1997	contact.gif

```
-rw-r--r-- nobody   5080 Mar 26  1997 dpcmast.gif
-rw-r--r-- nobody   570  Mar 26  1997 exit.gif
-rw-r--r-- nobody   587  Mar 26  1997 help.gif
-rw-r--r-- nobody   606  Mar 26  1997 order.gif
-rw-r--r-- nobody   592  Mar 26  1997 welcome.gif
```

```
/opt/interdoc/private/en/
```

permission	*user*	*size*	*date*	*name*	
-rw-r--r--	nobody	2356	Sep 19 17:48	etd_order.htm	- **Annex-XV**
-rw-r--r--	nobody	6823	May 19 19:97	t_scr2.htm	- **Annex-XVI**
-rw-r--r--	nobody	2710	Nov 24 18:32	t_scr3.htm	- **Annex-XVII**
-rw-r--r--	nobody	483	May 19 1997	t_scr4.htm	- **Annex-XVIII**

Field mapping

In order not to create new database fields in the DPC data schema and to include the job entry to appear in the Windows based program provided along with Interdoc system, the following fields of the Interdoc data space has been used to hold specific information required to process the job coming through this system.

Existing field	Modified content
paperStockType	payment method
documentPaperColor	name of card
paperType-size	card number
plex	card expiry date
specialInstruct	url and title

8.8 Interdoc web scripting

The core of this project is in establishing connection to the oracle database through the NCSA web server. Xerox interdoc™ is an interface cartridge which acts as a sandwich between the web page and the oracle database. As the NCSA server is configured to invoke "interdoc" for every page severed from the /opt/interdoc/.. directory, it is possible to embed interdoc specific scripts codes in a html formated file from that directory.

Interdoc web scripts are distinguished by <X-...> tags. When a page under the /pub or /mo (refer 8.6) is requested, the web server calls the "interdoc.csh" c-shell routine. This routine sets the ENV variables and passes the page to "interdoc" program. This interface cartridge parses the scripts and performs appropriate action over the Oracle (ver 7.1.6) RDBMS. The resulting output (if any) are embedded back into the html page itself. It is possible to call predefined procedures and also support some structured programming routines.

1)
```
<X-EXEC proc="accountQueryProc" data="aid=account_id">
```
This script line calls the accountQuery procedure with the above parameter.

2)
```
<X-IF ticket_id>
      <X-EXEC proc="jobQueryOneProc">
<X-ELSE>
      <X-EXEC proc="partQueryProc">
<X-/IF>
```
An if-then-else construction

3)
```
<!-- #VAR id_title="Place an Order">
```
Variable declaration

4) `<X-INC header.html>`

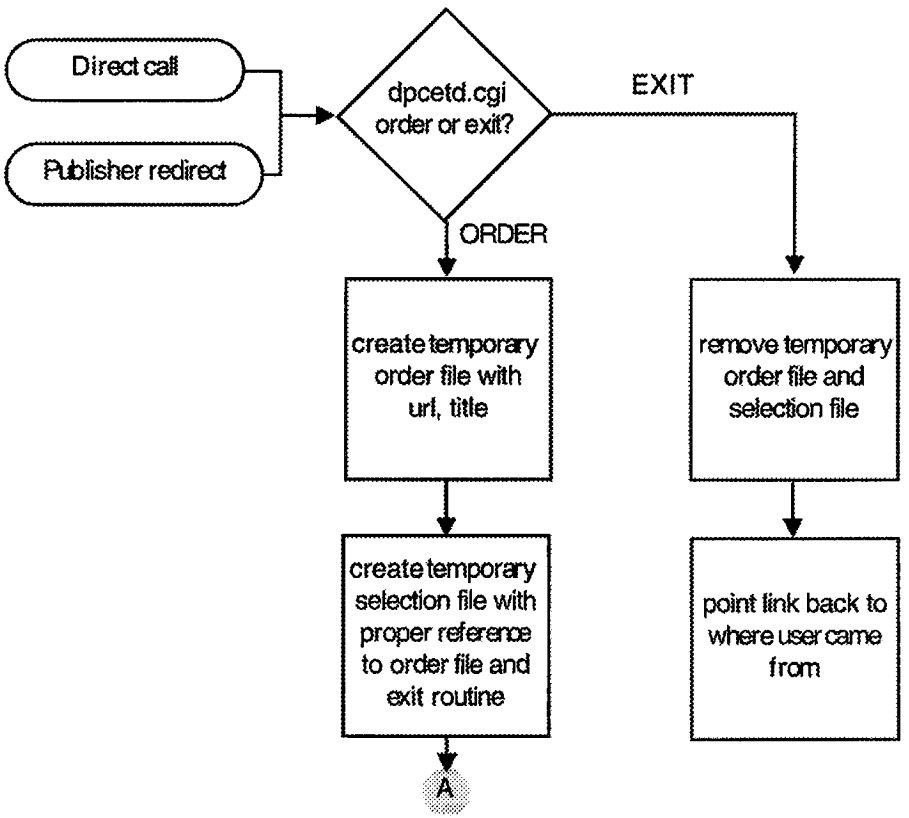
Including sub-script routines

5) `<!-- #config on=updateFlag set=form-->`

Conditional statement

8.9 Project process flow

External publisher's server invokes the dpcetd.cgi script on the DPC server. The process flow starts at this call and flows through as discussed below. After successful completion the exit screen takes the user back to their publisher's site.



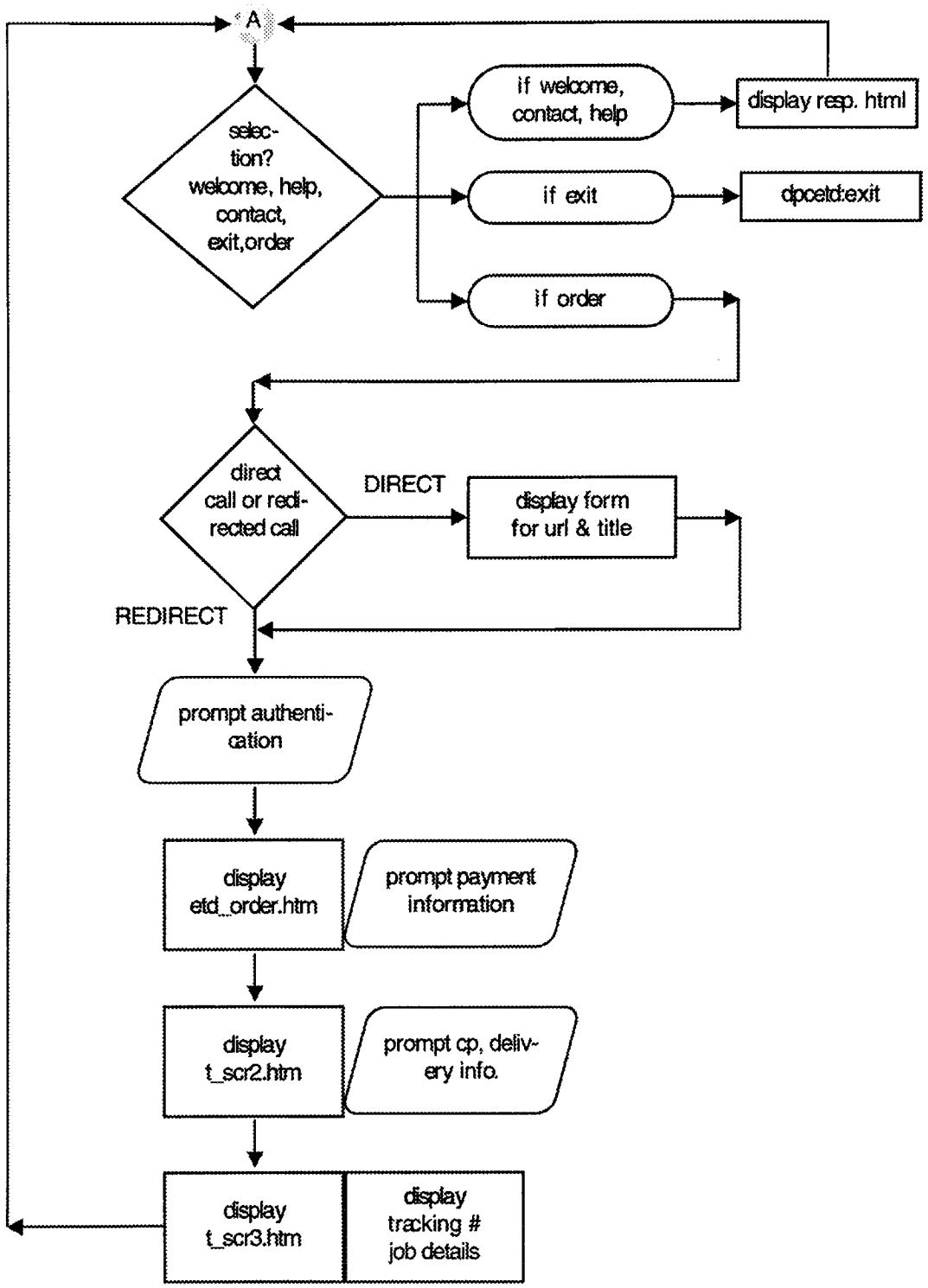


Figure 8.1, Project process flow

8.10 Testing

Test cases

To test all the functions of the system the following test cases are used:

1) check if "Request printed copy" marc-up is displayed in WML catalog

2) check if clicking the above hyperlink takes to DPC server and calls dpcetd.cgi

3) check if the url and title has been recognized by dpcetd script

4) check if the job is logged to DPC joblist

5) check if payment, delivery and special instruction fields has proper data

6) check direct call to DPC /etd site

7) check for url and title request when /etd site is reached directly

8) check the links to "shipping and handling", contact and help

9) check the execution of exit routine

10) check File Transfer module with authentication

Test results

Test 1 Using Netscape 4.0 browser, WML web site was accessed. From there, library catalog was clicked. Search by author was further selected. In the search field, "recene" was entered. On clicking the search button, the WML server returned the library catalog for James Recene's thesis. The marc field "Request printed copy of the thesis" was visible in the record. It was properly hyperlinked to dpcetd.cgi on DPC server with the GET syntax carrying the title and the url for the thesis pdf file.

Test 2 Clicking the "Request printed copy of the thesis" link displayed the welcome screen from the DPC /etd site, generated by the dpcetd.cgi script. In the selection.html frame, the filename for the temporary order file <tmp_order88309324929622.htm> was noted. The 14 digit random number that the temp file is created based on the time stamp and process id.

Test 3 The "Order" button was clicked. In the payment screen, the title of the thesis was displayed. The source of this screen was viewed. The value for "part_name" and "special_instruct" fields were checked. They reflected that of the current thesis title and url.

Test 4 After confirming payment and submitting delivery information, a job tracking number was displayed. Also, an entry was made in the DPC server.

Test 5 In the DPC server log, the details of this job number were checked. It contained all the information as described by the field mapping.

Test 6 Using the www browser, url http://129.21.203.108/etd was accessed. The static welcome screen was displayed.

Test 7 In the above state, "Order" button was clicked. It prompted for title and url. On entering the data, test case 3 was tested successfully.

Test 8 Shipping, contact and help links point to their respective html static pages. The information in these pages is wage and need to be rewritten by the DPC staff.

Test 9 From state test case 3, 4 and 7, "Exit" button was clicked. It displayed a thank you message and a link "BACK" pointing to http://albert.rit.edu/a?search=recene for test cases 3 and 4 and http://129.21.203.108/etd/index.html for test case 7. This proves that the return address is correctly reflected.

Test 10 From the Unix system prompt, the geturl.cgi perl script was run. The url, directory path, username and password were entered. The STDOUT is as follows:

```
%/opt/httpd/cgi-bin(2)> perl geturl.cgi
Enter URL, directory, username and password:
URL       :http://www.rit.edu/~mkd9472/log/log.html
Directory:/tmp
Username :mohan
Password :test
```

```
Checking connection... Successful
Please wait... URL is being fetched
The URL has been saved as /tmp/log.html
```

The /tmp/log.html was viewed with reflected the same file as in the original directory. A similar test was done for a binary file and it also proved success.

8.11 Business Model

The technical issues are ironed out in this project demo. The two other important issues are copyright issue and the royalty issue. With this new model of print distribution, these issues have to be considered outside the conventional boundary.

Copyright issue

The system captures the url of the authors' work on the Internet. If the author wants to protect his document, he can do so by using any of the Internet security system. The author discloses the security clearance to the printing partner. The author can then use standard "site log" utilities to keep track of the document access from the authorized printer. However, as like any digital system - illegal storage and redistribution of the content cannot be prevented. A legal agreement can be signed by the participating printer (much like the conventional printing projects) to protect the authors (publishers) interest.

Further reading on "Digital watermark" technology is recommended.

Royalty issues

Extending the tracking feature discussed above, the author can keep track of the number of access to his document by the printer. The royalty payment can be prearranged based on each access. Here again, if the printer prints extra copies illegally without disclosing to the author, a legal agreement could be of some protection. Another bottle neck will be to detect a complete file transfer from author to printer as against just a http hit.

A transaction system, linked to a credited financial institution can be developed as an extension to this system. This can automatically split the user's payment between the author, publisher and printer as pre-agreed.

8.12 Further work

Database of registered publishers

As discussed in section 8.3, a database module can be developed which generate publisher key id referencing other information like title, url, authentication, number of retrievals, contact etc.

In the dpcetd.cgi script, after the script checks for exit routine call, any external call can be passed which would return the url and title for the site to be logged. Upon return from the external module, the $userrequest::title and $userreuest::url variables must be set to the new value.

Credit card validation

When the user enters the credit card details, the standard validation algorithm can be incorporated (probably as Dynamic HTML scripting) to validate the information. There are a number of commercial software available for credit card validation as well. These could be used to run on the server along with the dpcetd script.

As described in the process flow diagram in section 8.9, the payment information is captured by the etd_order.htm script. This script by default updates the information to the database and calls t_scr2.htm. A run around can be done in etd_order.htm. Since, the database scripts are in the control of the "interdoc" interface cartridge, any processing must be done at the browser before the form is submitted calling t_scr2.htm. This can be done with any www browser supported scripting languages like JavaScript (or JScript).

Automatic file transfer

In this project model, the file transfer module is separated from the main application for the following reasons:

1) working over http connection on Internet documents, it is advisable to use a www browser to fetch the url supplied. This will provide access to view and print most of the documents like text, html and PDF

2) as we did not want to fetch the file at the time of user request but at the time of print processing, giving the control to the DPC staff

3) by using a browser to print the file, we use the browser's cache memory and avoid loading up the DPC server's file system

However, if for any reason these functions need to be automated as one step process it can be done as discussed further.

When the dpcetd.cgi script is called, it has the url. So, before the scripts starts to check for exit routine call, the file transfer module can be invoked with the necessary parameters. If the module return true, then further processing can be done if not the user can be informed of the error.

```
#Invoking geturl program

&ftp_error()
      unless (system("geturl.cgi $userrequest::url /repos"));
```

Chapter Nine

Summary and Conclusion

Summary

A general model has been satisfactorily developed and tested. The model addresses both the technology and the business issues facing the publishing industry in the wake of Internet.

The theoretical work related to developing this model, as included in this thesis, will provide a full understanding of the subject. The requirements of the publishing industry, the technology available today and the direction of the publishing and computer industry are all well discussed in this thesis.

The demonstration project allows a user coming to the WML to search for a thesis and request a printed copy. The delivery and payment information are registered in the DPC and the order is processed. Every module of the project is clearly described to facilitate understanding for those who might be interested in implementing this model in real world.

Conclusion

Internet is a promising media for doing publishing business. The explosive growth of the media is addressing a number of concerns pertaining to the publishing industry.

There is still a demand for printed materials inspite of the Internet revolution. It is not the technology, but the attitude of the reader that helps sustaining this demand.

Keeping up with the pace of technology, the model described in this thesis, has proven to be the ideal way to do online publishing business using Internet and on-demand printing technology.

Bibliography

Bibliography

Ames, Patrick. *Beyond paper: The official guide to Adobe Acrobat.* Adobe Press, 1993.

Copyright Clearance Center. World Wide Web site http://www.ccc.org/

Fenton, Howard M. and Romano, Frank J. *On-demand printing: the revolution in digital and customized printing.* Graphic arts technical Foundation, 1995.

Houldsworth, Andy. *Tuning your Images for the Web.* The seybold Report on Internet Publishing, Volume 1, Number 11, pp. 12-16. Seybold Publications, 1997.

Iversen, Osmund. *Server Push or Client Pull.* The seybold Report on Internet Publishing, Volume 1, Number 2, pp. 39-40. Seybold Publications, 1996.

Kinsgley, Lawrence. *Comdex Foretells of Web Printing.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 17-21. Seybold Publications, 1996.

Kinsley, Lawrence. *Consumer Behavior in Online Publishing.* The seybold Report on Internet Publishing, Volume 1, Number 5, pp. 9-14. Seybold Publications, 1997.

McCalpin, William. *Why Every Electronic Print Architecture is Wrong.* The Xplorer, Fall 93, p. 4. Xplor International, 1993

McKenzie, Matt. *Copyright Protection: Understanding Your Options.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 6-14. Seybold Publications, 1996.

McKenzie, Matt and Votsch, Victor. *Publishing Alternatives to the Web.* The Seybold Report on Internet Publishing, Vol 1, No 3, pp. 3-7. Seybold Publications, 1996.

Miller, Thomas E. *The 1997 American Internet User Survey.* Emerging Technologies Research Group, 1997.
World Wide Web site http://etrg.findsvp.com.

Romano, Frank J. *Digital media: publishing technologies for the 21st century.* Micro Pub. Press, 1996.

Walter, Mark. *W3C publishes draft of simplified SGML.* The seybold Report on Internet Publishing, Volume 1, Number 4, pp. 3-5. Seybold Publications, 1996.

Weinberger, Dr. David. One-to-One Publishing. The Xplorer, Fall 95, p.8. Xplor International, 1995.

Distribute, Then Print: Global Networks Take Demand Printing to remote Site. Seybold Report on Publishing Systems, Volume 24, Number 22. Seybold Publications, 1996.

GVU's 7th WWW User Survey. Graphics, Visualization & Usability Center, Georgia Tech Research Corporation, 1997.
World Wide Web site http://www.cc.gatech.edu/gvu/.

Internet is making a big impact on high-volume printing. The Seybold Report on Internet Publishing, Vol 1, No 3, pp. 32. Seybold Publications, 1996.

Microsoft pushes Channel Definition Format. The seybold Report on Internet Publishing, Volume 1, Number 8, pp. 25. Seybold Publications, 1997.

On-Demand Binding of Books. Seybold Report on Publishing Systems, Volume 24, Number 4. Seybold Publications, 1996.

Orlando Exhibit Review. The Xplorer, January 1997 Edition. Xplor International, 1997. World Wide Web site http://www.xplor.org/

Appendices

Appendix-A

Marc code of the WML catalog

Library Catalog

```
001     36294331
008     970130s1996     xx a      b     000 0 eng dntmIa
024     Color-printing|xData processing
040     RVE|cRVE
049     RVEL
090     Z286.E43|bR433 1996
100 1   Recene, Ronald James
245 13 An investigation into the use of the World Wide Web as an
       interface for distributing electronic documents to and
       from a remote digital color printing site /|cby Ronald
       James Recene
260     |c1996
300     1 v. (various foliations) :|bill. ;|c29 cm
500     Typescript
502     Thesis (M.S.)--Rochester Institute of Technology, 1996
504     Includes bibliographical references
520     ABSTRACT:  The World Wide Web and Internet are the most
       talked-about and fastest-growing mediums for information
       and electronic document distribution. Their growth has,
       and will continue to have, a great impact on all forms of
       media, due to their potential to reach millions of
       individuals. This project demonstrates the capabilities of
       the World Wide Web to perform, not only as a publishing
       vehicle, but as a means for communication and document
       distribution to a digital color printing facility.
520     In order to show this, a Web site was built that
       incorporated the utilities needed for the successful
       exchange of data, such as links to additional software
       applications available on the Web, downloadable ICC Color
       Management profiles of the digital color press, a
```

hypertext job estimate/information form, an uploadable FTP server, and directions on how to use the service and create the appropriate files. The result is a functional Web-based printing facility that eliminates the restrictions associated with geographical boundaries. The test to see if this site functioned properly was the successful implementation of the aforementioned applications and tools to create actual documents. Those documents, when put through the developed workflow, must exhibit the designers' original intent when reproduced on a remote digital press and when compared to their originals reproduced on that same press. The written portion of this thesis documents the procedures and rationale behind the methodology used.

650 0 Electronics in printing

650 0 Color-printing|xData processing

650 0 Facsimile transmission

650 0 Image transmission

650 0 Electronic publishing

650 0 World Wide Web (Information retrieval system)

856 7 |uhttp://design.rit.edu/etd/thesis001.pdf|zView thesis (pdf file)

856 7 |uhttp://129.21.203.108/cgi-bin/dpcetd.cgi?title=An+investigation+into+the+use+of+the+World+Wide+Web+as+an+interface+for+distributing+electronic+documents+to+and+from+a+remote+digital+color+printing+site&url=http%3A%2F%2Fdesign.rit.edu%2Fetd%2Fthesis001.pdf|zRequest printed copy of this thesis

Appendix-B

Runtime output of geturl.cgi

Error: <401> Unauthorized

```
[Mohan@q662q01]:/opt/httpd/cgi-bin(2)> perl geturl.cgi

Enter URL, directory, username and password:
URL       :http://www.rit.edu/~mkd9472/log/log.html
Directory:/tmp
Username :mohan
Password :guest

Checking connection... Error: <401> Unauthorized
```

Error: <404> Not found

```
[Mohan@q662q01]:/opt/httpd/cgi-bin(4)> perl geturl.cgi

Enter URL, directory, username and password:
URL       :http://www.rit.edu/unkown.txt
Directory:/tmp
Username :
Password :

Checking connection... Error: <404> Not found
```

Successfull fetching

```
[Mohan@q662q01]:/opt/httpd/cgi-bin(5)> perl geturl.cgi

Enter URL, directory, username and password:
URL       :http://www.rit.edu/~mkd9472/log/log.html
Directory:/tmp
Username :mohan
Password :test

Checking connection... Successful

Please wait... URL is being fetched

The URL has been saved as /tmp/log.html
```

Appendix-C

geturl.cgi

```
#!/usr/local/bin/perl
#Program to demonstrate File transfer and authentication
#
#Module as part of the Thesis work by Mohan Dhandapani
#Dec-23-97
#
#Uses LWP module
#Makes http connection to fetch the http url
#Uses username and password authentication is provided
#

  use LWP::UserAgent;
  $ua = new LWP::UserAgent;


  if ($#ARGV < 1) {
    #If command line arguments are not present prompt for them
    print "Enter URL, directory, username and password:\n";
    print "URL       :";
    chop($url = <stdin>);
    print "Directory:";
    chop($tempdir = <stdin>);
    print "Username :";
    chop($username = <stdin>);
    print "Password :";
    chop($password = <stdin>);
    print "\n";


#DEBUG
#     $url = 'http://www.rit.edu/~mkd9472/log/log.html';
#     $tempdir = '/tmp';
```

```
#     $username = 'mohan';
#     $password = 'test';
  }
  else {
    #Assign variables from command line arguments
    $url = $ARGV[0];
    $tempdir = $ARGV[1];
    if($ARGV[2]) {$username = $ARGV[2];}
    if($ARGV[3]) {$password = $ARGV[3];}
  }

  #Create temp file name on local
  @temp = split(/\//, $url);
  $tempfile = $tempdir."/".$temp[$#temp];
#DEBUG
#  print $tempfile;


  #Create request object
  $req = new HTTP::Request GET => $url;
  $req->authorization_basic($username, $password);
  if (! open(TEST,">$tempfile")) {
    print "Couldn't open $tempfile for writting.\n";
    exit;
  }

  #Open http connection
  $res = $ua->request($req);
  print "Checking connection... ";
  if ($res->is_success()) {print "Successful\n";}
  if ($res->is_error()) {
    print "Error: <".$res->code."> ".$res->message."\n";
    exit;
  }


  #Fetch the document and save to local
  print "Please wait... URL is being fetched\n";
```

```
    print TEST $ua->request($req)->content;
    close(TEST);
    print "The URL has been saved as $tempfile\n\n";


1;
#End of module
```

Appendix-D

dpcetd.cgi

```
#!/usr/local/bin/perl
## Script for processing PRINTED copy request for Elec. Thesis
## through WALLY.RIT.EDU

use CGI;
$query = new CGI;
$query->import_names('userrequest');


#Variable declaration
$secure_path              = '/opt/interdoc/private/en';
$etd_path                 = '/opt/httpd/htdocs/etd';
$order_file               = $secure_path."/etd_order.htm";
$template_index_file      = $etd_path."/template_index.html";
$template_selection_file = $etd_path."/template_selection.html";
$tmp_order_file           = "tmp_order".time()."$$.htm";
$tmp_selection_file       = "tmp_selection".time()."$$.htm";



#call exit routine if this script is called by EXIT button
&exit_routine unless (!$userrequest::remove);



#Open and read lines from order_file in secure path
if (! open(O_FILE,"$order_file")) {
      print "Content-type: text/plain\n\n";
      print "Couldn't open $order_file.\n";
       print "Print Request couldn't be processed.\n";
      exit;
}
```

```
@O_LINES=<O_FILE>;
close(O_FILE);

&clean_up;

#Massage Order file and write to a tmp file in secure directory
if (! open(UPDATE,">$secure_path/$tmp_order_file")) {
      print "Content-type: text/plain\n\n";
      print "Couldn't open $secure_path/$tmp_order_file";
      print "for writting.\n";
      print "Print Request couldn't be processed.\n";
      exit;
}
foreach (@O_LINES) {
   if (/!- perljobtitle ->/) {
      print UPDATE "Title: <B>$userrequest::title</B><P>";
      if ($ENV{'HTTP_REFERER'}) {
         ($u1,$temp) =  \
           ($ENV{'HTTP_REFERER'} =~ /([a-zA-Z:\/]*[\w.]*)/);
         print UPDATE "From: $u1\/<P>\n";
      }
   }
   elsif (/!- perljoburl ->/) {
      print UPDATE "<!- perljoburl ->URL  : ";
      print UPDATE "<A HREF=\"$userrequest::url\"> ";
      print UPDATE "$userrequest::url</A><P>\n";
   }
   elsif (/!- perlsplinstr ->/) {
      print UPDATE "<!- perlsplinstr ->";
      print UPDATE "<INPUT TYPE=hidden ';
      print UPDATE "NAME=\"special_instruct\" ";
      print UPDATE "VALUE=\"$userrequest::url\">\n";
   }
   elsif (/!- perlpartname ->/) {
      print UPDATE "<!- perlpartname ->";
      print UPDATE "<INPUT TYPE=hidden NAME=\"part_name\" ";
      print UPDATE "TYPE=text SIZE=40
VALUE=\"$userrequest::title\">\n";
```

```
    }
    else {
        print UPDATE $_;
    }
}
close (UPDATE);



#Read template selection file, massage and write a tmp file in
etd path
if (! open(S_FILE,"$template_selection_file")) {
      print "Content-type: text/plain\n\n";
      print "Couldn't open $template_selection_file.\n";
      print "Print Request couldn't be processed.\n";
      exit;
}
@S_LINES=<S_FILE>;
close(S_FILE);
if (! open(S_FILE,">$etd_path/$tmp_selection_file")) {
      print "Content-type: text/plain\n\n";
      print "Couldn't open $template_selection_file for ";
      print "writing.\n";
      print "Print Request couldn't be processed.\n";
      exit;
}
foreach (@S_LINES) {
   if (/<!-tmp_order_file->/) {
       $_=~ s/<!-tmp_order_file-
>/http:\/\/129.21.203.108\/mo\/$tmp_order_file/;}
   if (/<!-remove_tmp_order_file->/) {
       $_=~ s/<!-remove_tmp_order_file->/$tmp_order_file/;}
   if (/<!-return_to->/) {
       if ($ENV{'HTTP_REFERER'} eq
"http://129.21.203.108/etd/order.html") {
         $return_url = "http://129.21.203.108/etd/index.html";
       }
       else {$return_url = $ENV{'HTTP_REFERER'};}
```

```
        $_=~ s/<!-return_to->/$return_url/;
    }
    print S_FILE;
}
close(S_FILE);


#Updating Index file and displaying it to Browser
if (! open(I_FILE,"$template_index_file")) {
       print "Content-type: text/html\n\n";
       print "Couldn't open $indexfile . ";
       print "Print Request couldn't be processed.\n";
       exit;
}
@I_LINES=<I_FILE>;
close(I_FILE);

print "content-type: text/html\n\n";

foreach (@I_LINES) {
    if (/<!-tmp_selection_file->/) {
        $_ =~ s/<!-tmp_selection_file-
>/http:\/\/129.21.203.108\/etd\/$tmp_selection_file/;
    }
    if (/<!-tmp_order_file->/) {
        $_ =~ s/<!-tmp_order_file-
>/http:\/\/129.21.203.108\/mo\/$tmp_order_file/;
    }
    print $_;
}



#Cleanup - remove exisitng tmp files in private/en directory
#Will be done when User exits the Site by clicking on "Exit"
Button
#also when the program is called first time
#
```

```
sub clean_up {
opendir(DIR,$secure_path);
@tmpfilenames = grep (/tmp_[a-z0-9]+.htm/,readdir(DIR));
foreach (@tmpfilenames) {
#Debug    print "$_\n";
   system("/usr/bin/rm $secure_path/$_");
}
closedir(DIR);
opendir(DIR,$etd_path);
@tmpfilenames = grep (/tmp_[a-z0-9]+.htm/,readdir(DIR));
foreach (@tmpfilenames) {
#Debug    print "$_\n";
   system("/usr/bin/rm $etd_path/$_");
}
closedir(DIR);
}

#Exit routine calls clean_up and sends a Thank-You screen
#with BACK pointer to original site
#
sub exit_routine {
   &clean_up;
   print <<EOF;
content-type: text/html

<HTML><BODY BGCOLOR="#ffffff">
<CENTER>
Thanks for Stopping by our On-line Print Shop.<P>
<BR><P>
<A HREF="$userrequest::return" TARGET=_parent>BACK</A>
</BODY></HTML>
EOF
   exit;
}
1;
```

Appendix-E

contact.html

```
<HTML>
<HEAD>
<TITLE>
DPCETD:Contact Information
</TITLE>
</HEAD>

<BODY BGCOLOR="#ffffff">
<CENTER><IMG SRC="pix/dpcmast.gif" WIDTH=277 HEIGHT=58 BORDER=0
ALT="Digital Publishing Center at SPMS, RIT"><P></CENTER>


<H2>Contact:</H2>
<P>

<BLOCKQUOTE>
Barbara Birkett, Associate Professor<BR>
E-mail: babppr@rit.edu
<P>
Frank Cost, Assistant Dean<BR>
E-mail: fjcppr@rit.edu
<P>
Mohan Dhandapani, Graduate Student '97<BR>
E-mail: mkd9472@rit.edu
<P>
</BLOCKQUOTE>

</BODY>
</HTML>
```

Appendix-F

exit.html

```
<HTML><BODY BGCOLOR="#ffffff">
<CENTER>
Thanks for Stopping by our On-line Print Shop.<P>
<BR><P>
</BODY></HTML>
```

The dpcetd.cgi script generates an exit.html page on the fly with hyperlink pointing to the url from where the user came to the DPC site.

Hence the above static page is not used.

Appendix-G

help.html

```
<HTML>
<HEAD>
<TITLE>
DPCETD: Help
</TITLE>
</HEAD>

<BODY BGCOLOR="#ffffff">
<CENTER><IMG SRC="pix/dpcmast.gif" WIDTH=277 HEIGHT=58 BORDER=0
ALT="Digital Publishing Center at SPMS, RIT"><P></CENTER>

Help instructions to be updated.

</BODY>
</HTML>
```

Appendix-H

index.html

```
<HTML><HEAD>
<TITLE>DPCETD - Print Service</TITLE>
</HEAD>

<FRAMESET Rows="90,*" BORDER="0">
      <FRAME Name="mast" SRC="mast.html" Scrolling="No">
      <FRAMESET Cols="110,*" BORDER="0">
            <FRAME Name="selection" SRC="selection.html"
            Scrolling="Auto">
            <FRAME Name="board" SRC="welcome.html"
            Scrolling="Auto">
      </FRAMESET>

<NOFRAMES>
<BODY>
Frame Based Browser is required for the time being.<P>
Sorry for the inconvenience.<P>

</BODY>
</NOFRAMES>
</FRAMESET>

</HTML>
```

Appendix-I

mast.html

```
<HTML><HEAD>
<TITLE>DPCETD - Mast Head</TITLE>
</HEAD>
<BODY BACKGROUND="pix/bkg.jpg" BGCOLOR="#FFFFFF">
<CENTER><IMG SRC="pix/dpcmast.gif" WIDTH=277 HEIGHT=58 BORDER=0
ALT="Digital Publishing Center at SPMS, RIT"><P></CENTER>
</BODY>
</HTML>
```

Appendix-J

order.html

```
<HTML>
<HEAD>
<TITLE>
DPCETD: Order Form
</TITLE>

<BODY BGCOLOR="#ffffff">

<FORM ACTION="http://129.21.203.108/cgi-bin/dpcetd.cgi"
METHOD=Get TARGET=_top>
<TABLE BGCOLOR="dedeff" CELLSPACING=0 CELLPADDING=4 BORDER=0>
<TR>
<TD>
Enter the Title of the Electronic Document
</TD>
<TD>
<INPUT TYPE=Text NAME="title" VALUE="" SIZE=30>
</TD>
</TR>

<TR>
<TD>
Enter the URL of the Electronic Document
</TD>
<TD>
<INPUT TYPE=Text NAME="url" VALUE="" SIZE=30>
</TD>
</TR>

<TR>
<TD COLSPAN=2 ALIGN=Center>
<INPUT TYPE=Submit NAME="Place Order" VALUE="Submit">
```

```
</TD>
</TR>

</TABLE>
</FORM>

</BODY>
</HTML>
```

Appendix-K

selection.html

```
<HTML><HEAD>
<TITLE>DPCETC- Selection</TITLE>
</HEAD>
<BODY BACKGROUND="pix/bkg.jpg" BGCOLOR="#FFFFFF">
<CENTER>
<A HREF="welcome.html" TARGET=board><IMG SRC="pix/welcome.gif"
WIDTH=80 HEIGHT=25 BORDER=0 ALT="Welcome"></A><P>

<A HREF="order.html" TARGET=board><IMG SRC="pix/order.gif"
WIDTH=80 HEIGHT=25 BORDER=0 ALT="Place Order"></A><P>

<A HREF="help.html" TARGET=help><IMG SRC="pix/help.gif" WIDTH=80
HEIGHT=25 BORDER=0 ALT="Help"></A><P>

<A HREF="contact.html" TARGET=contact><IMG SRC="pix/contact.gif"
WIDTH=80 HEIGHT=25 BORDER=0 ALT="Contact"></A><P>

<A HREF="exit.html" TARGET=board><IMG SRC="pix/exit.gif" WIDTH=80
HEIGHT=25 BORDER=0 ALT="Exit DPC"></A><P>

</CENTER>
</BODY>
</HTML>
```

Appendix-L

template_index.html

```
<HTML><HEAD>
<TITLE>DPCETD - Print Service</TITLE>
</HEAD>

<FRAMESET Rows="90,*" BORDER="0">
      <FRAME Name="mast"
          SRC="http://129.21.203.108/etd/mast.html"
          Scrolling="No">
      <FRAMESET Cols="110,*" BORDER="0">
            <FRAME Name="selection"
                  SRC="<!-tmp_selection_file->"
                  Scrolling="Auto">
            <FRAME Name="board"
                  SRC="/etd/welcome.html"
                  Scrolling="Auto">
      </FRAMESET>

<NOFRAMES>
<BODY>
Frame Based Browser is required for the time being.<P>
Sorry for the inconvenience.<P>
</BODY>
</NOFRAMES>

</FRAMESET>

</HTML>
```

Appendix-M

template_selection.html

```
<HTML><HEAD>
<TITLE>DPCETC- Selection</TITLE>
</HEAD>

<BODY BACKGROUND="pix/bkg.jpg" BGCOLOR="#FFFFFF">
<CENTER>

<A HREF="welcome.html" TARGET=board>
<IMG SRC="pix/welcome.gif" WIDTH=80 HEIGHT=25 BORDER=0
ALT="Welcome">
</A><P>

<A HREF="<!-tmp_order_file->" TARGET=board>
<IMG SRC="pix/order.gif" WIDTH=80 HEIGHT=25 BORDER=0
ALT="Place Order">
</A><P>

<A HREF="help.html" TARGET=help>
<IMG SRC="pix/help.gif" WIDTH=80 HEIGHT=25 BORDER=0
ALT="Help">
</A><P>

<A HREF="contact.html" TARGET=contact>
<IMG SRC="pix/contact.gif" WIDTH=80 HEIGHT=25 BORDER=0
ALT="Contact">
</A><P>

<FORM ACTION="/cgi-bin/dpcetd.cgi" METHOD=Get TARGET=board>
<INPUT TYPE=HIDDEN NAME="remove"
VALUE="<!-remove_tmp_order_file->">
<INPUT TYPE=HIDDEN NAME="return" VALUE="<!-return_to->">
<INPUT TYPE=Image SRC="pix/exit.gif" BORDER=0>
```

```
</FORM>


</CENTER>
</BODY>
</HTML>
```

Appendix-N

welcome.html

```
<HTML>
<HEAD>
<TITLE>DPC - Welcome</TITLE>
</HEAD>

<BODY BGCOLOR="#ffffff">

<B>Welcome to the Digital Publishing Center at SPMS, RIT.</B><P>
Please click on 'Order' to process your print request.<P>
When prompted for Username and Password, you may enter your
personal ID established with DPC or enter with our guest/guest
ID.<P>

<B>Note:</B><P>
<UL>
<LI>If you were directed to this site by your publisher, the
source of your title is already registered by the system. Just
fill up the payment and delivery information.
<LI>If you entered this site directly, please enter the 'Title'
and 'URL' you like to be printed, followed by payment and
delivery information.
</UL>


</BODY>
</HTML>
```

Appendix-O

etd_order.html

```
<!-X-//-> $Workfile:   t_scr1.htm  $ $Revision:   1.17  $
<!-X-//-> Copyright (c) 1996, 1997 DPC, SPMS.  All Rights
Reserved.
<HTML>
<HEAD>
<TITLE>$ps_name; - On-Demand Printing</TITLE>
<!-#VAR id_title="On-Demand Printing"->
</HEAD>

<BODY BGCOLOR="#ffffff">

<FORM ACTION="/mo/t_scr2.html" METHOD=POST>
<X-//>BEGIN REQUIRED FOR INTERDOC
      <INPUT TYPE=hidden NAME=id_form VALUE="$id_document;">
      <INPUT TYPE=hidden NAME=id_db VALUE="partCreateProc">
      <INPUT TYPE=hidden NAME=id_cols VALUE="form=id_form,
message=part_name, part_name, documentPaperColors, orientation,
paperType_size, paperType_preFinish, paperType_opacity, plex,
finishingType, paperStockType, collation, edgeStitching,
slipsheeting, binding">
<X-//>END REQUIRED FOR INTERDOC


On-Demand Digital print request for<P>
<!- perljobtitle -><P>


Your request for the above title will cost <B>50 US Dollars</B>
plus <A HREF="ship_hand.htm" TARGET=Ship>shipping and
handling</A> costs.<P>

<TABLE BGCOLOR="#dfdfff" CELLSPACING=0 CELLPADDING=4 BORDER=0>
```

```
<TR><TD COLSPAN=2>
Enter your payment information:<P>
</TD></TR>

<TR><TD COLSPAN=2>
<B>Payment Method</B><BR>
<INPUT TYPE="Radio" NAME="paperStockType" VALUE="MasterCard"
CHECKED> Master Card
<INPUT TYPE="Radio" NAME="paperStockType" VALUE="VisaCard"> Visa
<INPUT TYPE="Radio" NAME="paperStockType" VALUE="Cash-Check">
Cash/Check on Pick-Up
<P>
</TD></TR>

<TR><TD COLSPAN=2>
<B>Name on Card</B><BR>
<INPUT TYPE="Text" NAME="documentPaperColors" VALUE="" SIZE=50
MAXLENGTH=32>
<P>
</TD></TR>

<TR><TD>
<B>Card Number</B><BR>
<INPUT TYPE="Text" NAME="paperType_size" VALUE="" SIZE=20
MAXLENGTH=32>
<P>
</TD>
<TD>
<B>Card Exp. Date</B> (mm/dd/yy)<BR>
<INPUT TYPE="Text" NAME="plex" VALUE="" SIZE=20 MAXLENGTH=32>
<P>
</TD></TR>

<TR><TD BGCOLOR="#eeeeee" COLSPAN=2>
<!- perlpartname -><INPUT TYPE=hidden NAME="part_name" TYPE=text
SIZE=40 VALUE="">
<INPUT TYPE=hidden NAME="orientation"          VALUE="Portrait">
<INPUT TYPE=hidden NAME="finishingType"       VALUE="tape">
```

```
<!- perlsplinstr -><INPUT TYPE=hidden NAME="special_instruct"
VALUE="">
<INPUT NAME="submit" TYPE=submit VALUE="Confirm Request">
</FORM>
</TD></TR>
</TABLE>

<P><HR><P>
<CENTER>
<A HREF="../pub/comment.htm" TARGET=New">Feedback</A><P>
<A HREF="/pub/copyrght.html">Copyright $copyMark; 1996
$ps_name;.</A>  All Rights Reserved.<P>
</CENTER>

</BODY>
</HTML>
```

Appendix-P

t_scr2.html

```
<!-X-//-> $Workfile:   torder.htm  $ $Revision:   1.12  $
<!-X-//-> Copyright (c) 1995, 1996 Xerox Corporation.  All
Rights Reserved.
<X-EXEC proc="accountQueryProc" data="aid=account_id, full=true">
<X-IF id_cols>                        <X-//> Add finishing
parameters
      <X-IF LITERALLY finishingType="SinglePortrait">
            <X-VAR edgeStitching=finishingType>
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="">
      <X-ELSEIF LITERALLY finishingType="SingleLandscape">
            <X-VAR edgeStitching=finishingType>
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="">
      <X-ELSEIF LITERALLY finishingType="DualLandscape">
            <X-VAR edgeStitching=finishingType>
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="">
      <X-ELSEIF LITERALLY finishingType="slipSheeting">
            <X-VAR edgeStitching="">
            <X-VAR slipSheeting=finishingType>
            <X-VAR collation="collated">
            <X-VAR binding="">
      <X-ELSEIF LITERALLY finishingType="uncollated">
            <X-VAR edgeStitching="">
            <X-VAR slipSheeting="">
            <X-VAR collation="uncollated">
            <X-VAR binding="">
      <X-ELSEIF LITERALLY finishingType="tape">
```

```
            <X-VAR edgeStitching="">
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="left">
      <X-ELSEIF LITERALLY finishingType="saddle">
            <X-VAR edgeStitching="">
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="">
            <X-VAR bookletMaker="true">
      <X-ELSE>
            <X-VAR edgeStitching="">
            <X-VAR slipSheeting="">
            <X-VAR collation="collated">
            <X-VAR binding="">
      <X-/IF>                          <X-//> Done adding finishing
parameters
                                 <X-//> Add stock parameters
      <X-IF LITERALLY paperStockType="Standard & Drilled">
            <X-VAR paperType_preFinish="PlainDrilled 3 0">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="Standard">
            <X-VAR paperType_preFinish="Plain 0 0">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="Fullcut & Drilled">
            <X-VAR paperType_preFinish="DrilledPreCutTab 3 0">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="Fullcut">
            <X-VAR paperType_preFinish="PreCutTab 0 0">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="PrecutTab & Drilled">
            <X-VAR paperType_preFinish="DrilledPreCutTab 3 3">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="PrecutTab">
            <X-VAR paperType_preFinish="PreCutTab 0 3">
            <X-VAR paperType_opacity="opaque">
      <X-ELSEIF LITERALLY paperStockType="Transparency &
Drilled">
```

```
            <X-VAR paperType_preFinish="PlainDrilled 3 0">
            <X-VAR paperType_opacity="transparency">
            <X-VAR documentPaperColors="noColor" ifset=force>
      <X-ELSEIF LITERALLY paperStockType="Transparency">
            <X-VAR paperType_preFinish="Plain 0 0">
            <X-VAR paperType_opacity="transparency">
            <X-VAR documentPaperColors="noColor" ifset=force>
      <X-ELSE>
            <X-VAR paperType_preFinish="Custom 0 0">
            <X-VAR paperType_opacity="opaque">
      <X-/IF>                        <X-//> Done* adding stock
parameters
      <X-VAR new_cols=id_cols oper=EXPAND>
      <X-EXEC proc=id_db data="files=, attr_list=new_cols">
<X-/IF>
<HTML>
<HEAD>
<TITLE>$ps_name; - Place an Order</TITLE>
<!-#VAR id_title="Place an Order"->
</HEAD>

<BODY BGCOLOR="#ffffff">

<!-X-//->Handle Errors
<X-IF id_date_error>
      <H2>$id_date_error;</H2>
      The <B>Required by</B> date you have entered is
unacceptable.  The supported date formats are:
      <UL>
      <LI><B>mm/dd/yy</B> or <B>mm-dd-yy</B> (11/25/95 or 11-25-
95)
      <LI><B>mm/dd</B> or <B>mm-dd</B> (11/25 or 11-25)
      <LI><B>month dd, yyyy</B> (November 25, 1995)
      <LI><B>month dd</B> (November 25)
      <LI><B>yyyy-mm-dd</B> (1995-11-25)
      <LI><B>dd. month yyyy</B> (25. November 1995)
      <LI><B>dd month yyyy</B> (25 November 1995)
      <LI><B>dd-mon-yyyy</B> (11-Nov-1995)
```

```
      <LI><B>dd-mon-yy</B> (11-Nov-95)
      </UL>
      Please check the date you have entered and use one of the
supported date formats.
      <P>
      <!—#config on=updateFlag set=form—>
      <HR>
      <P>
<X-ELSEIF id_entry_error>
      <H2>$id_entry_error;</H2>
      An invalid value has been entered into one of the fields.
Either:
      <UL>
      <LI>The copy count you have entered is invalid.  Copy
counts are always positive numeric values. Or
      <LI>The  instructions field has grown too large.  This
field is limited to 250 characters.
      </UL>
      <P>
      Please enter a valid value.
      <P>
      <!—#config on=updateFlag set=form—>
      <HR>
      <P>
<X-/IF>

Complete these ordering instructions and select <B>Submit
Order</B> when you are ready to submit the job.  If you can't
find the options you want, please describe what you would like
in the <B>Special Instructions</B> section.
<BR>

<FORM ACTION="/mo/t_scr3.html" METHOD=POST>
<X-//>BEGIN REQUIRED FOR INTERDOC
      <INPUT TYPE=hidden NAME=id_form VALUE="t_scr2.htm">
      <INPUT TYPE=hidden NAME=part_id VALUE="$part_id;">
      <INPUT TYPE=hidden NAME=part_name VALUE="$part_name;">
```

```
      <INPUT TYPE=hidden NAME=id_order_cols
VALUE="customer_status=new, private_status=new, copies, title,
required_by, delivery_method, purchase_order, delivery_to,
street1, street2, city, state_region, postal_code, country,
phone, fax, special_instruct">
<X-//>END REQUIRED FOR INTERDOC
      <INPUT NAME="title" TYPE=hidden VALUE="$part_name;">
$paperStockType;


<PRE>
Title:       $part_name;
Copies:          <INPUT NAME="copies" TYPE=text VALUE="1" SIZE=5>
<INPUT NAME="required_by" TYPE=hidden VALUE="09/09/99"
SIZE=15><INPUT TYPE=hidden NAME="purchase_order" VALUE=""
SIZE=25>

<INPUT TYPE=Hidden NAME="deliveryToMenu" VALUE="other">
Delivery:    <SELECT NAME="delivery_method">
<OPTION VALUE="Messenger">Via Messenger
<OPTION VALUE="Fedx">Via Fed Ex
<OPTION VALUE="UPS">Via UPS
<OPTION VALUE="FirstclassMail" SELECTED>Via FirstClassMail
<OPTION VALUE="Pickup">Via Customer Pickup
</SELECT>
<INPUT TYPE=hidden NAME="special_instruct"
VALUE="$special_instruct;">

Delivery Address:
Name:        <INPUT TYPE=text NAME="other_delivery_to" VALUE=""
SIZE=50>
Email:       <INPUT NAME="other_street1" TYPE=text VALUE=""
MAXLENGTH=256 SIZE=50>
Street:      <INPUT NAME="other_street2" TYPE=text VALUE=""
MAXLENGTH=256 SIZE=50>
City/State/Zip/Country:
```

```
            <INPUT NAME="other_city" TYPE=text VALUE=""
MAXLENGTH=256 SIZE=25>, <INPUT NAME="other_state_region"
TYPE=text VALUE="" MAXLENGTH=256 SIZE=5><INPUT
NAME="other_postal_code" TYPE=text VALUE="" MAXLENGTH=256
SIZE=10><INPUT NAME="other_country" TYPE=text VALUE=""
MAXLENGTH=256 SIZE=5>
Phone:       <INPUT NAME="other_phone" TYPE=text VALUE=""
MAXLENGTH=256 SIZE=21> Fax: <INPUT NAME="other_fax" TYPE=text
VALUE="" MAXLENGTH=256 SIZE=21>
</PRE>

<INPUT NAME="submit" TYPE=submit VALUE="Submit Order">
<INPUT NAME="reset" TYPE=reset VALUE="Reset to Defaults">
</FORM>

<P><HR><P>
<CENTER>
<A HREF="../pub/comment.htm" TARGET=New">Feedback</A><P>
<A HREF="/pub/copyrght.html">Copyright $copyMark; 1996
$ps_name;.</A>  All Rights Reserved.<P>
</CENTER>

</BODY>
</HTML>
```

Appendix-Q

t_scr3.html

```
<!—X-//—> $Workfile:   treceipt.htm  $ $Revision:   1.14  $
<!—X-//—> Copyright (c) 1995, 1996 Xerox Corporation.  All
Rights Reserved.
<X-IF id_order_cols>
      <X-VAR id_address_flag=true><X-INCLUDE custdl.htm>
      <X-VAR new_order_cols=id_order_cols oper=EXPAND>
      <X-EXEC proc=jobCreateProc data="attr_list=new_order_cols">
      <X-EXEC proc=jobAddPartProc data="tid=ticket_id,
pid=part_id">
<X-ELSEIF NOT ticket_id>            <X-//>for uploaded single
form
      <X-EXEC proc=partQueryOneProc data="pid=part_id,
full=true">
<X-/IF>

<HTML>
<HEAD>
<TITLE>$ps_name; - Thank you</TITLE>
<!—#VAR id_title="Thank you"—>
</HEAD>

<BODY BGCOLOR="#ffffff">

<X-EXEC proc="jobQueryOneProc" data="tid=ticket_id, full=true"
result=db>
Thank you for placing your order.<P>

Your Job tracking number is <B>$db.ticket_id;</B>,  with the
following details.<P>

For further assistance you may contact the DPC service personal
by Telephone at (716)475-5793 or through <A
```

```
HREF="../pub/comments.htm" TARGET=New>Feedback Form</A>. Please
quote the Job Ticket number for reference.<P>

<HR><P>

<PRE>
Job Title:          $db.title;
Submit Date:              $db.ticket_date;
Job Tracking Number:      $db.ticket_id;

Number of Copies:  $db.copies;

Credit Card Type:          $db.paperStockType;
Name on Credit card:       $db.documentPaperColors;
Credit Card #:             $db.paperType_size;
Card Expiry Date:          $db.plex;

Delivery Method:           $delivery_method;
Deliver to:                 Name       - $other_delivery_to;
                             Email      - $other_street1;
                             Address    - $other_street2;
                                          $other_city;
                                          $other_state_region;
                                          $other_postal_code;
                                          $other_country;
                             Tel        - $other_phone;
                             Fax        - $other_fax;

</PRE>
<BR>

<!-
<FORM ACTION="/mo/t_scr4.htm" METHOD=POST
ENCTYPE="multipart/form-data">
<X-//>BEGIN REQUIRED FOR INTERDOC
<INPUT NAME="part_id" VALUE="$part_id;" TYPE=hidden>
<INPUT NAME="ticket_id" VALUE="$ticket_id;" TYPE=hidden>
<INPUT NAME="id_files" VALUE="sourceFiles" TYPE=hidden>
```

```
<INPUT NAME="id_db" VALUE="partAddFilesProc" TYPE=hidden>
<INPUT NAME="id_url" VALUE="receipt.html" TYPE=hidden>
<INPUT NAME="id_label" VALUE="Confirm Order" TYPE=hidden>
<X-//>END REQUIRED FOR INTERDOC

<INPUT NAME="submit" TYPE=submit VALUE="Confirm your Order">
</FORM>
—>
<!— ask user to exit —>Please EXIT to return.



<P><HR><P>
<CENTER>
<A HREF="../pub/comment.htm" TARGET=New">Feedback</A><P>
<A HREF="/pub/copyrght.html">Copyright $copyMark; 1996
$ps_name;.</A>  All Rights Reserved.<P>
</CENTER>

</BODY>
</HTML>
```

Appendix-R

t_scr4.html

```
<!—X-//—> $Workfile:   treceipt.htm  $ $Revision:   1.14  $
<!—X-//—> Copyright (c) 1995, 1996 Xerox Corporation.  All
Rights Reserved.
<HTML>
<HEAD>
<TITLE>$ps_name; - Thankyou message</TITLE>
<!—#VAR id_title="Thank you"—>
</HEAD>

<BODY BGCOLOR="#ffffff">

<FONT SIZE=+2>
Thank you for placing your on-line print request with DPC. Your
order will be processed in 3 working days and sent to you/made
available for collection.<P>
</FONT>


<X-INC footer.html>
</BODY>
</HTML>
```