Network Administration and Security in a Digital Library Environment

By

JVM.Joseph, E.Soundararajan, C.Jayakumar and S.Venkadesan

Library & Information Services

Indira Gandhi Centre for Atomic Research

Kalpakkam – 603 102

Tamilnadu, INDIA

URL: www.igcar.ernet.in

Email: joe@igcar.ernet.in

ABSTRACT

The contribution of networking technologies for a library is significant. The need for proper administration and security for the library LAN is important as considered against the contents. This paper discusses about the methods and issues of network administration in a library followed by the security methods to be adapted on various levels in order to protect the contents from hackers/unauthorized access. The usage of data compression methods, network traffic management algorithms, encryption methods, metadata standards, and backup and recovery procedures are highlighted.

KEYWORDS: Network security, Firewall, Quality of Service, Metadata, Network Traffic, Cryptography, Content Security, Data Compression, ATM, Digital content preservation

0. INTRODUCTION

Digital Library systems are the computerized information storage and retrieval systems connected to computer networks and have maximum flexibility, durability, immunity, stability, scalability, accessibility and compatibility with other media at minimum cost, space and maintenance. Information dissemination of catalogue, abstract and full text formats over the network in digital format are the basic functions of a digital library apart from acquiring, digitizing and archiving of contents. The networking of library resources means that patrons can access the information from their desktops irrespective of their physical location. Digital Library makes the library without walls. This necessitates the management of library Network and securing of contents.

1. LIBRARY LAN & INTERNET

For a digital library to be useful, there must be a user community and a means for those users to reach the library. The simplest, closest, and most restrictive access is via a desktop station connected directly to the server on which the digital library collection is held. This limits use to the in-house users and constitutes a model which is simple to maintain, but not necessarily desirable in this age of networked information.

The most closed access in practice is across a Local Area Network (LAN) serving the organization’s campus. This provides remote access but within the geographical limits of the organization. To make the remote access possible, Internet technology can be used. Publishing of contents using World Wide Web (WWW) is a simpler method for remote access.

1.1 System Requirements & Content Creation

Two forms of systems need to be considered; hardware and software. While the software provides the functionality to make the digital library work, the hardware provides the underlying resources and processing.

Hardware Systems	Software Systems
Server / Desktop Systems	Server Class Operating System
Capturing Devices	Desktop Operating System
LANs	Programming / Scripting Language
Routers/ Switches	Web/ FTP/ Mail Server Softwares
Modem/ISDN	Content Creation Software
RAID arrays	Data Base Management System
Uninterrupted Power Supply (UPS)	Firewall & Protection Softwares
Tape/ Dick Backups	Office Automation Package
Printers	

Table 1- System Requirements for a Digital Library

Content Creation is one of the basic functions in a digital library. In- house documents of a library can be digitized using scanners and classified using metadata descriptive standards like XML / MARC and should be archived for the access by patrons.

2. NETWORK ADMINISTRATION

Once a network has been installed and has been running for a while, users begin to expect a certain level of performance. As changes are made to the network, its performance might degrade, and this degradation is often gradual. The ability to troubleshoot network problems effectively is a culmination of skills acquired through professional training, self-study, and on-the-job experience. The factors that contribute to making a good network troubleshooter are effective tools, good documentation, and experience.

2.1 Traffic Management

Quality of service (QoS) allows network administrators to use their existing resources efficiently and to guarantee that critical applications receive high-quality service without having to expand as quickly, or even over-provision their networks. Deploying QoS means that network administrators can have better control over their networks, reduce costs, and improve customer satisfaction. Asynchronous Transfer Mode (ATM) Networks are thought to transmit data with varying characteristics. Different applications need various QoS. Some applications like telephony may be very sensitive to delay, but rather insensitive to loss, whereas others like compressed video are quite sensitive to loss. ATM QoS Categories include Constant Bit Rate (CRA), Variable Bit Rate (VRA), Available Bit Rate (ABA) and Unspecified Bit Rate (UBA). A Network Administrator can decide the traffic management using the above methods.

2.2 Server Management

Digital library consists of several servers like Email server, FTP Server, Web Server, Contents Server etc. These Servers need to be given higher priority in the network performance. Apart from tuning the servers from the Operating System and Database levels, it is required that these servers are connected into high performance network switches and the QoS should ensure that these critical servers receive high priority in network traffic management.

2.3 User Management

It is the responsibility of a network administrator to create and manage the user groups of a library. Granting privileges based on the user groups, authentications based on user ID password, domain and work group control, Virtual LAN (V-LAN) management are some of the important user management activities that a network administrator should be familiar with.

3. NETWORK SECURITY

One of the critical and challenging tasks before the network administrator is ensuring the security of the library network, which means ensuring security of the contents. There will be several security threats like unauthorized access, hacking the web sites, executing the command illegally, data destruction, denial of service etc. Various levels of security measures would ensure the security of the network.

3.1 Firewall

In order to provide a level of separation between an organization's intranet and the Internet, *firewalls* have been employed. A firewall is simply a group of components that collectively form a barrier between two networks. The Firewall computer can reach the protected network and the Internet. This protected network cannot reach the Internet, and the Internet cannot reach the protected network. The simplest form of firewall is Linux machine with one network connection (an Ethernet card or modem) connected to the Internet and the other connected to the Gateway of a private network.

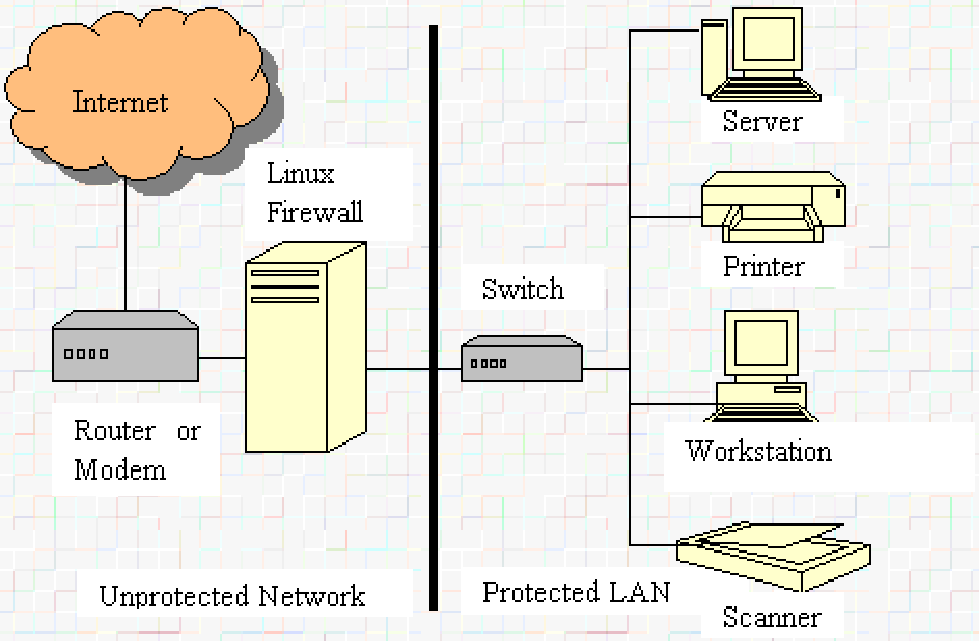

Figure 1 - A Sample Linux Firewall

Linux has a feature called IP Masquerading built in that can allow the operating system to act as a Network Address Translation (NAT) router and firewall for the entire network. Every Packet flowing to and from the Internet and the private network is filtered by the firewall. There are 3 basic type of firewall:

Ø Application Gateways

Ø Packet Filtering

Ø Hybrid Systems

Selecting an appropriate firewall for the library or for the Organisation is very important decision, which should be based on the security policies of the Library / Organisation, expert analysis and advice.

3.2 Operating System Level Security

Operating Systems are the main loopholes for the security threats. Tuning the operating system at various levels may be the task of system administrator, it should be done in coordination with network administrator. NT and Unix based operating systems like Solaris, Linux, UnixWare are designed to provide tighter security. For Linux operating system, a freeware would be an ideal choice for libraries. It provides lot of utilities and packages that could be useful for digital libraries. The Secure Shell System (SSH) of Linux provides encryption and authentication on connections. Encryption is using codes to protect the packets of data while in transit. Authentication is a process for verifying if a packet of data or a connection is valid. Linux log files are useful for tracing the individual connection attempts.

There are 2 levels of security which need to be provided:

3.2.1 Physical Security: The general rule for physical security is that if someone can get physical access to the hardware, he can gain access to all of the data and its trusted relationships with other machines on the network. Some of the guidelines for physical security are:

Ø Keep the servers in locked room with network and power cables snipped off.

Ø Disable booting from floppy drive and CD-ROM Drives as well.

Ø Protect BIOS settings with password

Ø Keep the Backup media in a safe place.

3.2.2 Securing Services: The services offered by the operating system in networked environments need to be protected. The system administrator should clearly configure each service in the view of security attacks. If some of the services are not required, then they should be disabled and passwords must be based on some mathematical algorithms, which will be difficult to break. File system security should be based on the user and group privileges. Excessive care must be taken while tuning the attack-pruned services like telnet, ftp. Services must not run with super user privileges (root user).

3.3 Database Level Security

Data base security is more concerned with data security and unauthorized access. Data base security is inherent part of the database design. Some of the data base security measures would be:

Data base Integrity

User Authentication

Access Control

Availability

Consistency

The database design should be based on these measures. MySQL database, a freeware database that comes with Linux makes use of the full security advantages of Linux.

4. CONTENT SECURITY

The ultimate security in library will mean only the security of contents. As the volume of information is growing rapidly, today's information era is measured in terms of giga bytes and terra bytes. Hence there is a need for a systematic approach to store, protect and retrieve the contents. Data compression methods and data encryption algorithms can be used for representing giga bytes of information. Access to the content should be modeled on the privilege levels. Data base management system could greatly help in managing giga bytes of contents. The measures related to content security in a library are

- Preservation of digital contents
- Intellectual Property Rights
- Authorized access
- Backup and recovery

4.1 Digital Content Authentication

Digital Content Security could be achieved through encryption methods. Encryption is the transformation of data into a form unreadable by anyone without a secret decryption key. Its purpose is to ensure privacy by keeping the information hidden from anyone for whom it is not intended, even those who can see the encrypted data. Authentication in a digital setting is a process whereby the receiver of a digital message can be confident of the identity of the sender and/or the integrity of the message. Authentication protocols can be based on either conventional secret-key cryptosystems like DES or on public-key systems like RSA authentication.

4.2 Retrieval

The success of a digital library system ultimately lies in how the users can access the information. Retrospective information access and interactive search facility are the great benefits of the digital contents which allow the users to search thousands of documents to the full text level at a time which might be impossible with the print materials. Relevance ranking of contents is an important issue which should ensure that the users get what they actually need in the first screen. There are several approaches to that like subject based thesaurus, hit count, frequency, currency etc. A simple web based interface for the retrieval of contents should be provided.

4.3 Backup & Recovery

Security of the data must be assured in both the immediate and longer terms. Immediate security can be provided by a Redundant Array of Inexpensive Disks (RAID) which spreads the data across a number of disks in a way that allows, even in case of a failure, system to still function while the failed component is replaced. In the longer term, regularly taking and checking backups of the contents ensures security. Backup policy should be well planned by the library so that in case of any disaster, the last backup should be able to restore the system as complete as possible. Some of the backup media are DVD ROM- R/RW, DVD-RAM, and DAT –TAPE.

4.4 Digital Content Preservation

Digital preservation means planning, resource allocation, and application of preservation methods and technologies necessary to ensure that digital information of continuing value remains accessible and usable. There could be several methods for digital preservation like preservation of the technology used, preservation of the physical media on which the data is held, migration and reformatting of the underlying data without change in its intellectual content etc. Issues relating to intellectual property rights and copyright will need to be considered when preserving digital materials for long term access.

5. CONCLUSION

Digital libraries should be well prepared to face the administrative and security challenges of their network and contents. To improve the technical knowledge and competency of the conventional librarian, training programme, technical courses and workshops would be helpful. It is necessary to have separate network administrator for the library. Library should take the expert advise for designing the network architecture, decide the security principles and content management issues. A thorough tuning of operating system, database management systems in view of security, preservation of digital contents, ensuring copyright protection, improved network bandwidth and performances are the main tasks in a digital library

REFERENCES

1) Marilyn Deegan, Simon Tanner. *"Digital Future : Strategies for the Information age*". Library Association Publishing, London 2002. pp178-208.

2) Venkadesan, S., Narayanan, A. "Digital Library initiative in an Indian Research Library: An experience Report". *Proceedings of the Conference on Recent Advances in Information Technology (READIT-99)*; October 28-29, 1999: 111-125.

3) Thomas Schenk et al. "*Red Hat Linux System Administration*". SAMS Techmedia, New Delhi 2000. pp739-817.

4) Matt Curtin. "*Introduction to Network Security*". March 1997.

URL: <http://www.interhack.net/pubs/network-security.pdf>

5) CERN Web maker. "*Data Base Security*". 1994.

URL: <http://www.cs.nps.navy.mil/curricula/tracks/security/notes/chap09_1.html>

6) Sun Micro Systems. *"The Digital Library Toolkit*". March 2000. URL: <http://www.sun.com/products-n-solutions/edu/whitepapers/pdf/digital_ library_toolkit.pdf>

7) Linux Firewall and Security Site. 2001. URL: <http://www.linux-firewall-tools.com/linux/?

8) Illinois Pulsar-Based Optical Interconnect (iPOINT), "*ATM Introduction*" 1998.

URL: <http://ipoint.vlsi.uiuc.edu/abr/atm_intro_frame.html>

-

BRIEF BIOGRAPHY OF AUTHORS

J.Veda Michael Joseph is the Network Administrator of Library & Information Services at Indira Gandhi Centre for Atomic Research. His subjects of interests are Discrete Mathematics, Networking, Operating Systems and Database Management. His current work includes Network administration, Web mining and Giga byte Information Retrieval and storage.

E. Soundararajan is the Systems Engineer of Library & Information Services at Indira Gandhi Centre for Atomic Research. His subjects of interests are Software engineering, Database Management and Operating System. His Current Work Includes System Administration, Software Development and providing technical solutions for Digital Library.

C. Jayakumar

S. Venkadesan