Basics of cyber infrastructure requirement for GIS, computing, storage, archive, network and firewall infrastructure in GIS lab

Dharmendra Kumar

Topics we will learn

- Introduction and requirement
- Cyber infrastructure for GIS
  - Software
  - Hardware
    - Peripherals and field kit
    - Miscellaneous
  - Data
- GIS lab
  - Computing
  - Storage
  - Archive
  - Network and
  - Firewall

Introduction

- Many GIS applications are implemented within organizations and have to play an important role.
- The effectiveness of GIS projects therefore depends as much on the futuristic planning of hardware and software requirements, as on technicalities of its implementation.
- It is therefore necessary to address the question of how GIS can be successful in an organization and continue to do so for many years before upgradations of hardware and software are needed.
- GIS and the host organisation have a close and two-way relationship.

Requirements

- There is a necessity to analyze the requirements keeping in view the following:
- Organization's goals and objectives
- Analysis of current practices and problems encountered and redundancies
- Organizational issues - management structure, resources, staff
- Training
- Applications
- Base map requirements
- Conceptual data base design
- System integration
- Cost of GIS implementation : applications development, data conversion, hardware, software, resources, training, update and maintenance
- The Strategic Implementation Plan over a period of 3-5 years

What is GIS Software?

- GIS software is a computer-based tool for examining geographic patterns, trends and relationships.
- They allow organizations to store, analyze and visualize data for geographic locations anywhere on Earth.
- By connecting geography with data, we understand data through a geographic context.
  - **GIS & Mapping Software**
  - **Imagery & Visualization Software**
  - **Data Analysis Software**
  - **Software Use Terms**

GIS & Mapping Software

- ArcGIS Desktop
- ArcGIS Pro
- Business Analyst Desktop
- ENVI
- GeoDa
- QGIS

Imagery & Visualization Software

- AutoCAD
- Drone2Map
- Gephi
- PhotoScan
- Rhino
- SolidWorks
- STK

Data Analysis Software

- Mathematica
- MATLAB
- NVIDIA CUDA Toolkit
- NVivo
- OpenRefine
- Processing
- R & RStudio
- Stata 16

List of other GIS Software

- ArcGIS Pro
- ArcGIS Desktop
- AutoCAD Map 3D
- Bentley Map
- FalconView
- FME® Feature Manipulation Engine
- GE Smallworld
- GeoDa
- Global Mapper
- Golden Software MapViewer
- QGIS 2
- SAGA GIS
- GRASS GIS
- gvSIG
- Hexagon Geomedia
- IDRISI TerrSet
- ILWIS
- Manifold GIS
- MapInfo Professional
- Maptitude
- OpenJump
- QGIS 3
- uDig
- WhiteBox GAT

Software Use Terms

- Use of licensed software is restricted to members of the community for the purposes of research, education, and scholarship. Under the licenses, users generally may not:
- Redistribute the software or materials or permit anyone other than a member of the community to use them
- Remove, obscure or modify any copyright or other notices included in the software or materials
- Use the software or materials for commercial purposes
- When a member of the organization terminates affiliation with the organization or a member's affiliation with the organization is terminated by the organization or a person not affiliated with the organization leaves campus, the member or person must destroy all personal copies of the software or derived data. Users are individually responsible for compliance with these terms.

Hardware Computers

- This is an area that changes rapidly, so think minimum to maximum with the maximum being best.
- Processor speed—The faster the better—800 MHz or faster Pentium® III or the "latest" speeds is best.
- Hard drive—20 GB or larger; if networked for data storage, the smaller amount may be okay.
- CD–ROM drive—24X or better.
- RAM—512 MB RAM minimum—512–1,024 MB is better.
- Internet access (ISDN, T1)—speed of downloads.
- CD–RW drive. Monitor: 17-inch monitor , 19- or 21-inch monitor is better.
- Graphics card—usually 32 MB graphics RAM or better. If using ArcGIS™ 3D Analyst ™, a true OpenGL card with 64 MB or better is required.

Hardware Computers

- Networked if using a server. Floppy drive. Microsoft® Office 2000, NT, or XP—Professional Edition.
- Anti-virus software.
- Laptops with two batteries.
- Transmitters and receivers for wireless capabilities (8011.a if possible).
- Cabinet for computers with plug-in recharges.
- Tables that lock together and are on wheels.
- Good chairs on rollers.
- Smart board.
- Overhead projector that can be swiveled in various directions.

Peripherals

- One color laser printer per eight to 10 computers with PostScript ® support
- One black-and-white printer per five computers with PostScript support
- Oversized plotter—needed only after program is going and really requires this as it is expensive
- Large format printers
- Digitizer—GTCO—12 x 12 is good for initial program, table later if necessary
- Scanner—flatbed 8 l/2" x 11" okay for initial program with scanner software
- Overhead projection system—necessary for demos of exercises and projects
- Laptop or PDAs—for field work; good if this is in long-term plan and field kit if required

Peripherals: Field kit

- The GIS department requires toolboxes that can be used for surveying in the field. Tool boxes may include:

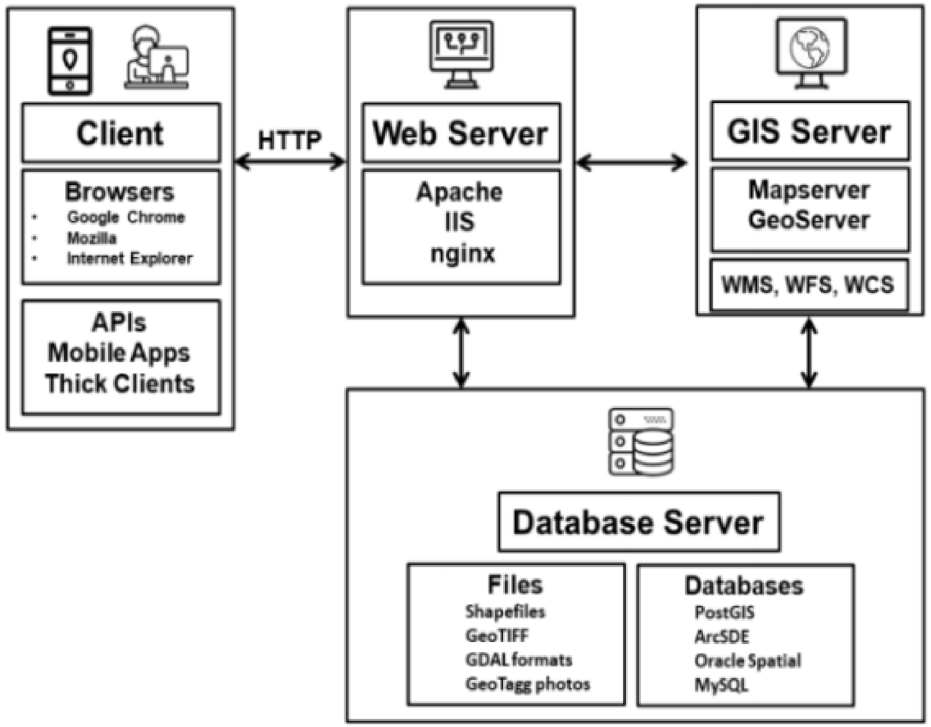
- Surveying tape
- Flash light
- Trowel
- Pencils
- Sharpee
- Duct tape
- Chaining Pins
- Wood Tape
- Tape Measure
- Electric tape
- Calculator
- Ruler
- Compass
- Hammer
- Whiskers
- Nails
- Spray paint
- Gloves
- Plumb- bob
- Rebar Caps

Miscellaneous

- Digital camera—great for fieldwork and student projects
- GPS equipment and software (differentially correcting is best) with compatible software Windows CE demo—use with ArcPad
- Extension power cords and cables
- Surge protectors
- Adequate power for all the system's needs
- Security for hardware and software

Recall: Topics we will learn

- GIS lab
- GIS lab assessment
- Advisory Committee
- Computing
- Storage
- Archive
- Network and
- Firewall



GIS lab assessment

- The first step in designing a needs assessment is to define the objectives, function, and design of the GIS program and the center or lab facility. Some important questions to ask are
  - Who are the users (faculty, staff, students) and what are their needs (occasional use, set class times per semester, free access at all times)?
  - Are you offering modules within courses, courses within existing degree programs, a degree or certificate program, or a mix of these offerings?
  - Will the GIS center or lab facility be a single or multipurpose facility?
  - Will the GIS center or lab facility be open or restricted to particular users at set times?
  - Will the computers be networked or stand alone?
  - How will software, computer, and facility security be provided?
  - What type of technical support exists on campus and how will technical support be provided to the center or lab facility?

GIS lab assessment: Advisory Committee

- **An advisory committee** : selected group of representatives from the university, local government agencies, school systems, private consulting firms, and other institutions that help provide a broad perspective on the goals, needs, and success of the GIS program.
  - This committee should be set up early in the planning process. Other subcommittees may also need to be formed from time to time to deal with specific tasks.
  - One of these may be a "technical committee" that has specific expertise in technical aspects of the program while another committee may be formed that specifically addresses the overall needs.
  - Members of the advisory committee should rotate every year or two to include representatives from other local businesses, industries, educational institutions, and government agencies.
  - This will generate new energy, enthusiasm, and ideas for long term success of the GIS program

Responsibilities and Decisions

- Some of the responsibilities and decisions the committee can assist with
  - Determining program needs
  - Determining local GIS needs
  - Designing curriculum to meet needs
  - Maintaining realistic timetable and goals
  - Determining available resources of each institution
  - Determining what equipment and facilities exist
  - Determining what faculty and support staff exist
  - Establishing an advertising and marketing campaign
  - Determining available funding
  - Supporting a GIS coordinator or manager

Computing

- The process of using computer technology to complete a given goal-oriented task.
- Geo-spatial Cyber infra

C. Yang et al. / Computers, Environment and Urban Systems 34 (2010) 264–277

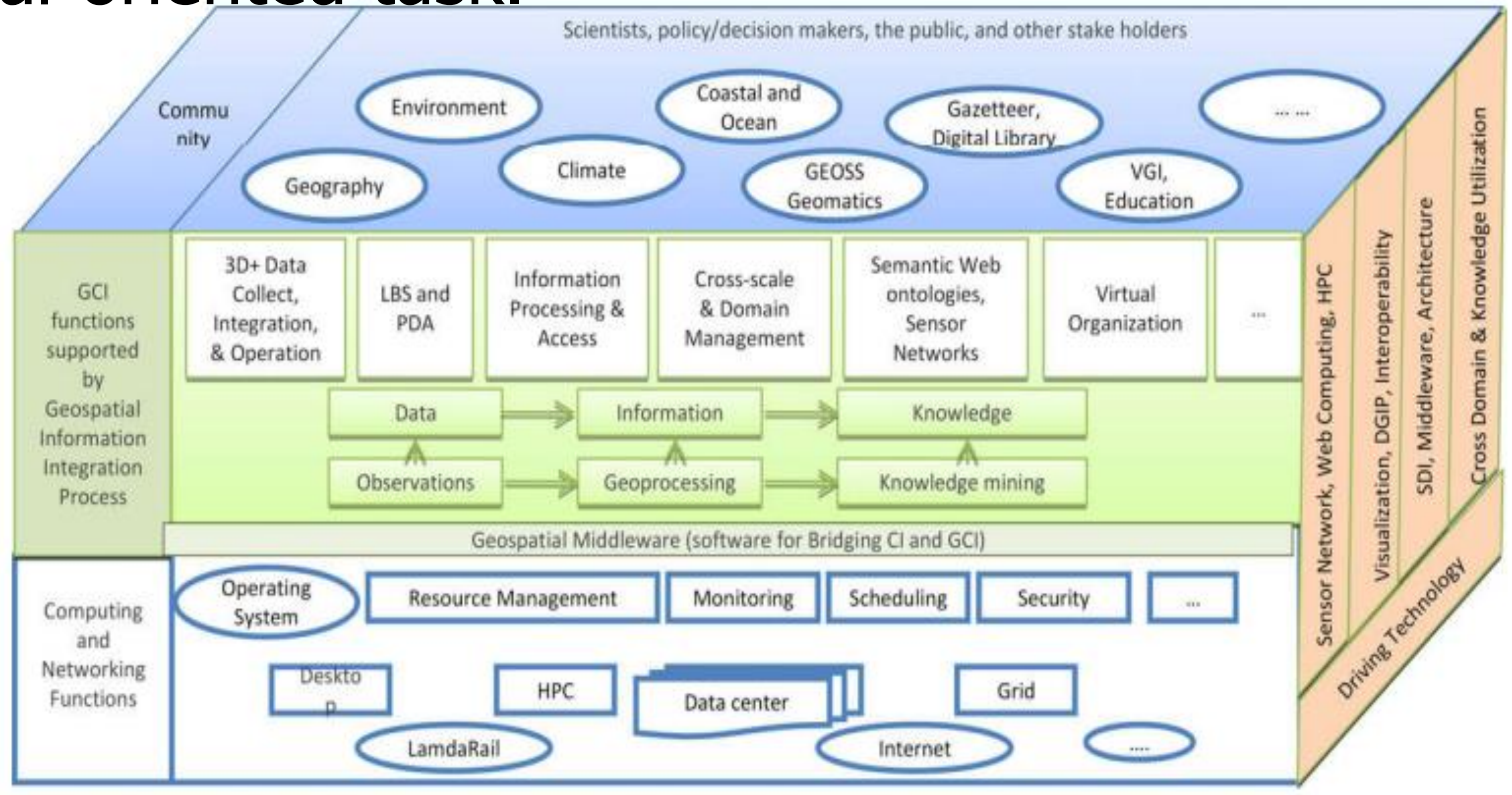


Fig. 2. GCI framework cube.

Storage

The three types of GIS Data are spatial, attribute, metadata

1.spatial data

1.vector data

1.Point Data — layers containing by points (or "events") described by x,y (lat,long; easting, northing)

2.Line/Polyline Data — layers that are described by x,y points (nodes, events) and lines (arcs) between points (line segments and polylines)

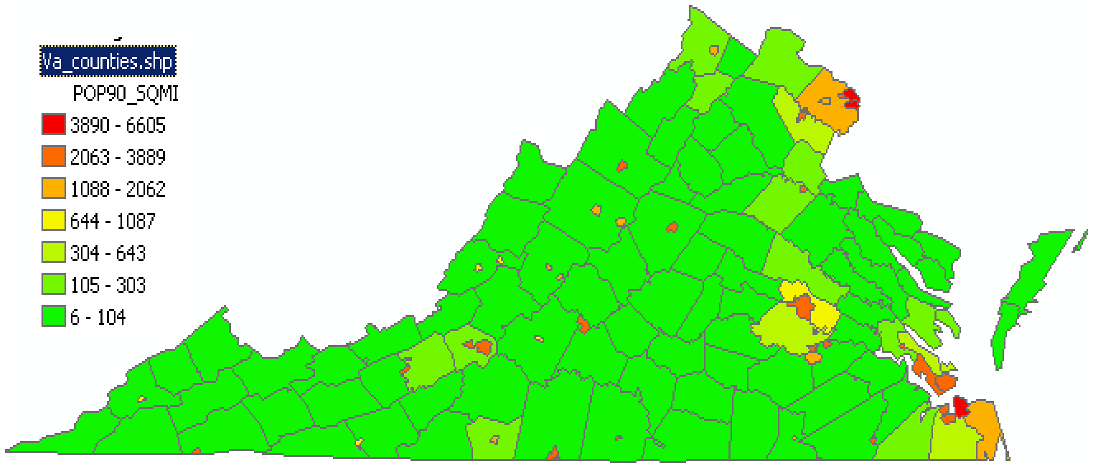
3.Polygon Data — layers of closed line segments enclosing areas that are described by attributes Polygon data can be "multipart" like the islands of the state of Hawaii.

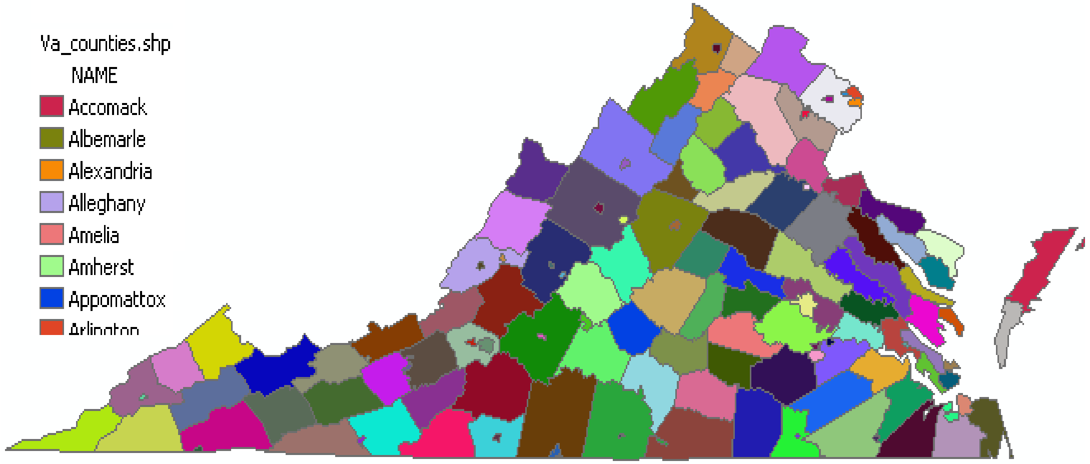
2.raster or grid data (matrices of numbers describing e.g., elevation, population, herbicide use, etc.

3.images or pictures such as remote sensing data or scans of maps or other photos. This is special "grid" where the number in each cell describes what color to paint or the spectral character of the image in that cell. (to be used, the "picture" must be placed on a coordinate system, or "rectified" or "georeferenced")

4.TINs – Triangular Irregular Networks – used to discretize continuous data

5.Terrain datasets built from lidar and other point clouds.





Storage

2.attribute data are non-spatial characteristics that are connected by tables to points, lines, “events” on lines, and polygons (and in some cases GRID cells)

1.A point, vector or raster geologic map might describe a “rock unit” on a map with a single number, letter or name, but the associated attribute table might have

- age
- lithology
- percent quartz
- etc, for each rock type on the map.

2.most GIS programs can either plot the polygon by the identifier or by one of the attributes

3.The above examples from the following project shows two ways to portray census data in Virginia. In the top image, each county/city gets a name and unique color , and in the bottom, the population density per square mile is read from the layer’s attribute table and plotted using a different color for each class of density.

3.Metadata

1.metadata are the most forgotten data type

2.absolutely necessary if you’re going to use data, or if someone is going to use your data later (or your derivative information)

3.contains information about

1.scale
2.accuracy
3.projection/datum
4.data source
5.manipulations
6.how to acquire data

4.many different “standards” for collection and presentation of metadata, such as FGDC used by US gov’t agencies.

Archive

Archiving involves moving data that is no longer frequently accessed off primary systems for long-term retention. Unlike data on primary storage, which needs to be frequently accessed and modified, archived data is retained for long periods of time, and best searched for when needed.

Loss of Data

Confused and Stressed Out Employees

Increased Backup Costs

Search Impact

Legal Compliance- industry requirement

Security- ransomware

GDPR and Discovery

- Lawfulness, fairness and transparency.
- **Purpose** limitation.
- Data minimization.
- **Accuracy**.
- Storage limitation.
- **Integrity** and **confidentiality** (**security**)
- Accountability.

Archive

- As an IT expert, we may search for data archiving solutions. The technology chosen must overcome a number of challenges.
  - **Growing volumes of data.** The sheer volume data produced by the average organization continues to grow exponentially. Data archiving solutions must be able to scale easily to manage ever-larger storage volumes without an increase in administrative time.
  - **Legal compliance.** Regulatory frameworks are constantly evolving and becoming more stringent. Data archiving strategies must simplify the task of archiving data to comply with a wide variety of regulations and make it easier and less time-consuming for administrators to fulfill legal inquiries and compliance reporting requests.
  - **Data security.** Data held in archive must be protected from loss, corruption, theft and compromise.
  - **Need for access.** Archived data may need to be accessed by a wide variety of users. Data archiving solutions can simplify access requirements with role-based permissions that allow users to easily search and retrieve without intervention by IT teams.
  - **Increased costs.** Data archiving solutions can help to reduce the cost of storage by archiving only essential data, by automatically eliminating data that no longer needs to be retained and by eliminating duplicated information.
  - **Management complexity.** Configuring data archiving and retention policies can be remarkably time-consuming. Superior data archiving technologies must minimize the administrative burden on IT teams for managing archived data.

Network

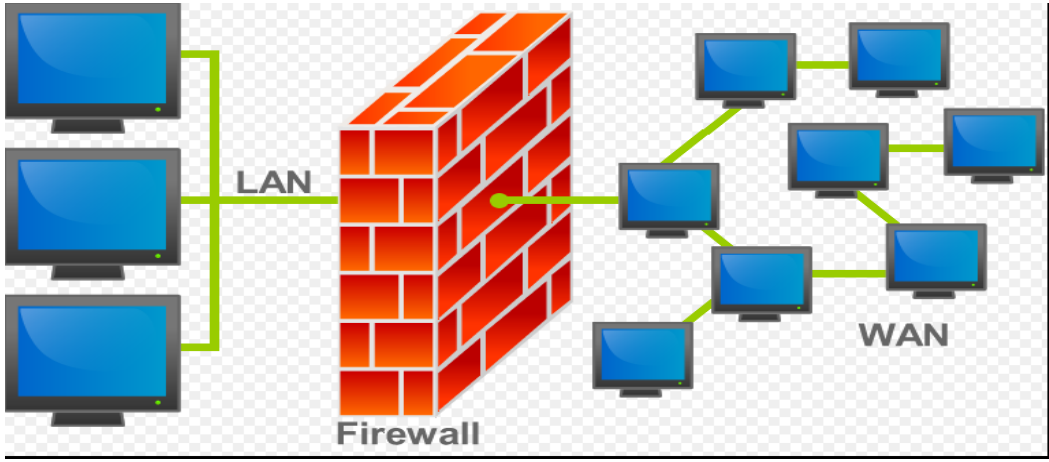
- **Network planning and design** is an iterative process, encompassing topological design, network-synthesis, and network-realization, and is aimed at ensuring that a new telecommunications network or service meets the needs of the subscriber and operator. The process can be tailored according to each new network or service.
- A traditional network planning methodology in the context of business decisions involves five layers of planning, namely:
  - need assessment and resource assessment
  - short-term network planning
  - IT resource
  - long-term and medium-term network planning
  - operations and maintenance.

Network

- The network planning process involves three main steps:
  - **Topological design**: This stage involves determining where to place the components and how to connect them. The (topological) optimization methods that can be used in this stage come from an area of mathematics called Graph Theory. These methods involve determining the costs of transmission and the cost of switching, and thereby determining the optimum connection matrix and location of switches and concentrators.
  - **Network-synthesis**: This stage involves determining the size of the components used, subject to performance criteria such as the Grade of Service (GOS). The method used is known as "Nonlinear Optimisation", and involves determining the topology, required GoS, cost of transmission, etc., and using this information to calculate a routing plan, and the size of the components.
  - **Network realization**: This stage involves determining how to meet capacity requirements, and ensure reliability within the network. The method used is known as "Multicommodity Flow Optimisation", and involves determining all information relating to demand, costs, and reliability, and then using this information to calculate an actual physical circuit plan.

Firewall

- In computing, a **firewall** is a network security system that monitors and controls incoming and outgoing network traffic based on predetermined security rules. A firewall typically establishes a barrier between a trusted network and an untrusted network, such as the Internet.



An **air gap**, **air wall**, **air gapping** or **disconnected network** is a network security measure employed on one or more computers to ensure that a secure computer network is physically isolated from unsecured networks, such as the public Internet or an unsecured local area network. It means a computer or network has no network interfaces connected to other networks, with a physical or conceptual air gap, analogous to the air gap used in plumbing to maintain water quality.

Firewall

Packet filter:

The first reported type of network firewall is called a packet filter, which inspect packets transferred between computers. The firewall maintains an access control list which dictates what packets will be looked at and what action should be applied, if any, with the default action set to silent discard. Three basic actions regarding the packet consist of a silent discard, discard with Internet Control Message Protocol or TCP reset response to the sender, and forward to the next hop. Packets may be filtered by source and destination IP addresses, protocol, source and destination ports. The bulk of Internet communication in 20th and early 21st century used either Transmission Control Protocol (TCP) or User Datagram Protocol (UDP) in conjunction with well-known ports, enabling firewalls of that era to distinguish between specific types of traffic such as web browsing, remote printing, email transmission, file transfer.

The first paper published on firewall technology was in 1987 when engineers from Digital Equipment Corporation (DEC) developed filter systems known as packet filter firewalls. At AT&T Bell Labs, Bill Cheswick and Steve Bellovin continued their research in packet filtering and developed a working model for their own company based on their original first-generation architecture.

Connection tracking

From 1989–1990, three colleagues from AT&T Bell Laboratories, Dave Presotto, Janardan Sharma, and Kshitij Nigam, developed the second generation of firewalls, calling them circuit-level gateways. Second-generation firewalls perform the work of their first-generation predecessors but also maintain knowledge of specific conversations between endpoints by remembering which port number the two IP addresses are using at layer 4 (transport layer) of the OSI model for their conversation, allowing examination of the overall exchange between the nodes.

Firewall

Application/ Layer 7

Marcus Ranum, Wei Xu, and Peter Churchyard released an application firewall known as Firewall Toolkit (FWTK) in October 1993. This became the basis for Gauntlet firewall at Trusted Information Systems.

The key benefit of application layer filtering is that it can understand certain applications and protocols such as File Transfer Protocol (FTP), Domain Name System (DNS), or Hypertext Transfer Protocol (HTTP). This allows it to identify unwanted applications or services using a non standard port or detect if an allowed protocol is being abused.

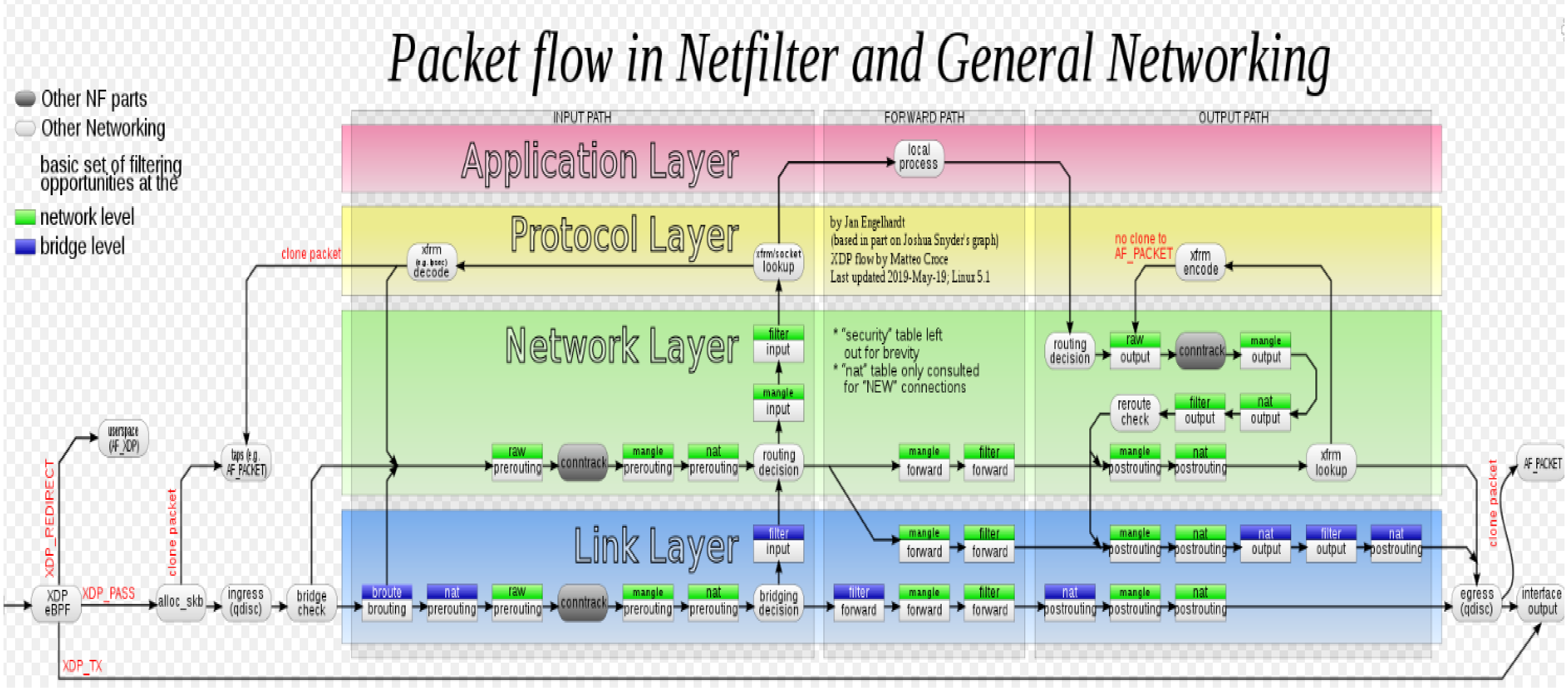
As of 2012, the next-generation firewall provides a wider range of inspection at the application layer, extending deep packet inspection functionality to include, but is not limited to:

- Web filtering
- Intrusion prevention systems
- User identity management
- Web application firewall

Endpoint specific

Endpoint based application firewalls function by determining whether a process should accept any given connection. Application firewalls filter connections by examining the process ID of data packets against a rule set for the local process involved in the data transmission. Application firewalls accomplish their function by hooking into socket calls to filter the connections between the application layer and the lower layers. Application firewalls that hook into socket calls are also referred to as socket filters.

Firewall



Packet flow in Netfilter and General Networking
Other NF parts
Other Networking
basic set of filtering opportunities at the
network level
bridge level
INPUT PATH
FORWARD PATH
OUTPUT PATH
Application Layer
Protocol Layer
Network Layer
Link Layer
local process
by Jan Engelhardt
(based in part on Joshua Snyder's graph)
XDP flow by Matteo Croce
Last updated 2019-May-19; Linux 5.1
* "security" table left out for brevity
* "nat" table only consulted for "NEW" connections
clone packet
no clone to AF_PACKET
xfrm decode
xfrm/socket lookup
xfrm encode
xfrm lookup
routing decision
reroute check
conntrack
bridging decision
bridge check
broute brouting
ingress (qdisc)
egress (qdisc)
interface output
alloc_skb
XDP eBPF
XDP_REDIRECT
XDP_PASS
XDP_TX
userspace (AF_XDP)
taps (e.g. AF_PACKET)
AF_PACKET
raw prerouting
mangle prerouting
nat prerouting
mangle input
filter input
mangle forward
filter forward
raw output
mangle output
nat output
filter output
mangle postrouting
nat postrouting

Cyber security

- Injection, Broken Authentication, Sensitive Data Exposure, XML External Entities (XXE), Broken Access Control, Security Misconfigurations, Cross Site Scripting (XSS),Insecure Deserialization, Using Components with Known Vulnerabilities, Insufficient Logging and Monitoring etc.
- Cyber security consists of a continuing cycle of structured actions to:
  - Identify (understand state and risks to systems, assets, data, and capabilities)
  - Protect (implement the appropriate safeguards)
  - Detect (implement ability to identify a Cybersecurity event)
  - Respond (implement ability to take action following a Cybersecurity event)
  - Recover (implement resilience and restoration of impaired capabilities)

Laboratory aspects

- Some method to protect software and hardware from theft or tampering. Note: This includes physical theft as well as digital tampering by students; remember that students will need to be able to save their work. Limit or track access, but students must still have access for saving projects and download.
- Most access questions come down to personnel, security, and time management issues. Without the personnel to be on hand for troubleshooting and security reasons, access to the facility will be affected. It may be necessary for lab managers or coordinators to play dual roles in order to provide more open lab schedules. If lab assistants are put in charge, policies should be established regarding access (who gets keys), shutting down computers and monitors, emergency contacts, and so forth.

Laboratory aspects

- The most important security issues deal with hardware theft, tampering with computers, backing up the database, and virus protection.
- These issues can be dealt with by having in place physical or digital security systems and following processes and procedures to be sure they are effectively used.
- Rooms housing GIS programs need to have some form of security including security for doors and windows as well as **access security systems** (keypads or other devices). Also needed is some form of security (locks or locked cabinets) for computers and other hardware within the room or facility.
- **Security for computer operating systems, programs, and files** are also vital. These can include efficient password policies, antivirus software, and backup and restore software and procedures to be sure these are used effectively. How many times have we heard or said the day a hard drive crashes or the system acquires a virus, "I was going to back up my files yesterday"?
- A reliable **backup system** for the GIS center or facility database should be a top priority in maintaining the GIS database. A dependable backup system will restore the backup database to another computer and be operational again in minutes. A schedule for regular backups depends on the size of the database and the number of changes being made to it. The GIS manager or lab technician should be responsible for these regular backups.

Laboratory aspects

- Computer viruses need to be viewed as a "when" not an "if" situation. Viruses and security issues go hand in hand. There are several excellent virus detection software programs available at very reasonable costs.
- University computer facilities most likely have site licenses for most of these programs of which university faculty and labs should be entitled.
- However, the software is only effective if continual virus updates are performed to keep the detection program current. This can be done directly through the Internet.
- Since it is difficult at times to keep up with the number of viruses being created, the best security is a system backup.
- In addition, policies should be established to limit the use of USB storage in the lab and to screen any USB storage before it is accessed by a program as well as limiting the downloading of material from the Internet.

DISCLAIMER

The material for the presentation has been compiled from various sources such as books, tutorials (online and offline), lecture notes, several resources available on Internet. The information contained in this lecture/presentation is for general information and education purpose only. While we endeavor to keep the information up to date and correct, we make no representation of any kind about the completeness and accuracy of the material. The information shared through this presentation material should be used for educational purpose only.