Communication Protocols
and Internet Architectures

Harvard University

Lecture #8

Instructor: Len Evenchik
cs40@evenchik.com or evenchik@fas.harvard.edu

ALIGHLSOD1701

Lecture Agenda

- Course Logistics
- Q&A and Topics from Last Week
- Connection Management, 5-tuple
- Network Address Translation (NAT)
- NAT Related Protocols: RFCs 3022, 4787, 5382 and 5389
- One Minute Wrap-Up

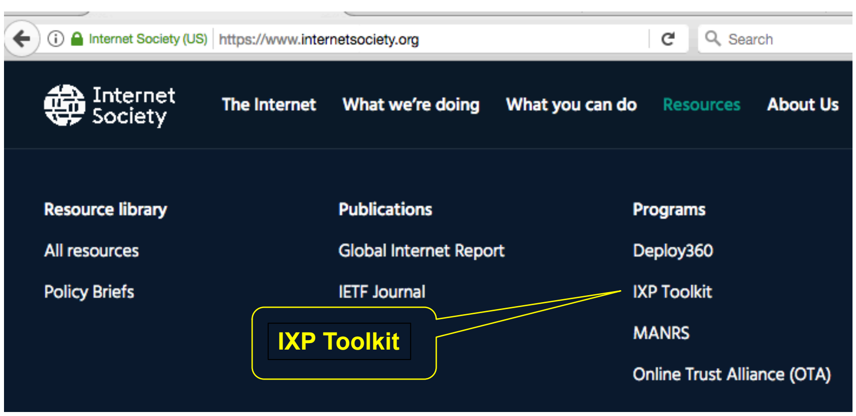
Course Logistics

Course Logistics

- Please contact the instructor if you will not be able to submit homework #3 when it is due. The solutions will be posted the day after the homework is due so that you have time to review the solutions before the midterm.
- MIDTERM EXAM – Please make sure to read all of the logistics and administrative information about the midterm that is in the weekly course information sheet. Please contact a TA or the instructor if you have any questions.
- The section meetings held right before the midterm will include a review for the exam.
- **Please submit a one minute wrap-up each week. Thank You!**

Q&A
Topics from Last Week

IXP Toolkit via the Internet Society
https://www.internetsociety.org



ALIGHLSOD1701

IXP Resources
https://ixptoolkit.org/

https://ixptoolkit.org/maps/resources/ Search

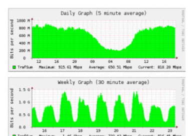
Resources

This "Resource Page" is meant to provide you with studies, reports, analyses, and information on IXPs, the business case for IXPs (more coming soon!), economic information, and other related data. Please send us resources that we have not included, and we will add them here: feedback@ixptoolkit.org

IXP Traffic Statistics, Data, and Links to Internet Data

There are excellent efforts underway to track IXP data to better understand traffic patterns, network health, IXP members' technical activities, and create a visible means for investors to see that there are users on networks – investment potential.

The Internet Society and others are encouraging more network measurement and traffic diagnostics.



Fortaleza IXP Traffic Data http://www.ptt.br/trafego/agregado/ce

ALIGHLSOD1701

IXP Toolkit Map
https://ixptoolkit.org/

https://ixptoolkit.org/map/ Search

IXP toolkit

IXPs | Maps and Data | Best Practices | People | Training

ALIGHI

SP3 Protocol Framework

- Service
  - The Service is a description of what the protocol does, not how it is done. This should be a few sentences long.
- Purpose
  - The Purpose describes the specific functionality that the protocol provides and how it is accomplished. Examples are flow control, error detection, error correction, etc.
- Packets
  - The Packet layout determines how the various bits and fields within the packet are defined, assembled and used.
- Procedures
  - The Procedures describe the various packet exchanges and the reason for each exchange.

UDP Pseudo Header

- UDP checksum provides end-to-end error detection, but not correction.
- UDP checksum includes a pseudo header which is conceptually prefixed to the UDP header
- Pseudo header includes the IP source and destination addresses, the IP protocol field, and the UDP length
- The checksum is very important, but it is also important to understand that its calculation violates the concept of strict protocol layering.

RFC 4614 (2006)
TCP is Complicated and Implementation Specific

- RFC 4614 was written to provide a roadmap to TCP. It describes the most important RFCs and just as important, it explains what prior work is no longer relevant.
- It is important to note that it was written in 2006. See RFC 6247.
- This roadmap is divided into four main section:
  - Basic (core) Functionality
  - Recommended Enhancements
  - Experimental Extensions
  - Historic Extensions

TCP Sequencing, Flow Control and Segmentation
Some General Characteristics

- A TCP data stream is a sequence of octets (bytes). This is different than the link layer protocols we have studied which keep track of entire frames (not bytes.)
- TCP uses a sliding window and this window identifies bytes not packets
- TCP does not know anything about the bytes it is sending. (As you would expect with transport protocols.)
- The size of the Flow Control window changes dynamically over time
- Retransmission timeout also changes dynamically (based on RTT) over time

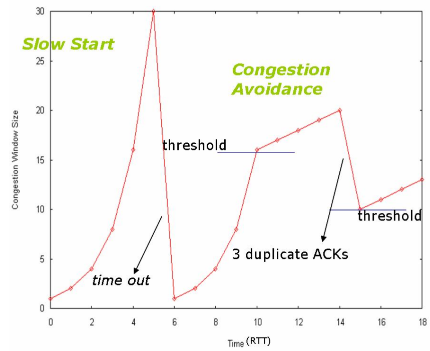
TCP Congestion Control (1)

- The assumption today is that networks are reliable (but of course, not perfect) and therefore packets are dropped due to network congestion.
- A second assumption is that network congestion occurs at routers (or other network devices) due to bursts of traffic.
- Congestion control and flow control are very different. (A classic drawing represents this with pipes and overflowing buckets of water.)
- Congestion control and Slow Start were not part of RFC 793. Van Jacobson designed them in 1988.
- TCP "Slow Start" addresses congestion control, not flow control.

TCP Congestion Control (2)

- Note that "Slow Start" is a misnomer. It is actually exponential growth.
- TCP uses four intertwined Congestion Control algorithms and mechanisms (RFC 5681)
  - slow start
  - congestion avoidance
  - fast retransmit
  - fast recovery.
- TCP Congestion Avoidance is additive-increase, multiplicative-decrease (AIMD)
- Finally, the first assumption we make for congestion control is that networks are reliable, and this does not apply to wireless networks. What does this mean for real world implementations?

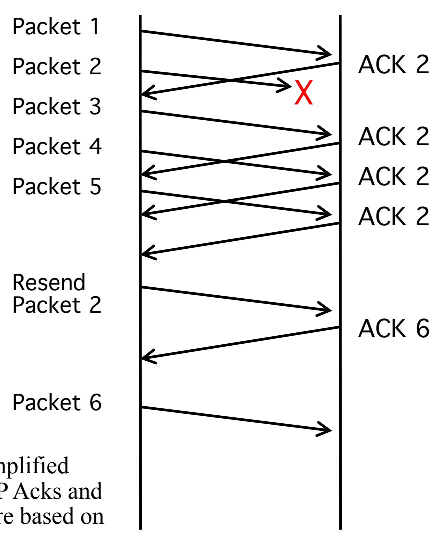
TCP Slow Start (3)
Fast Recovery and Fast Retransmit Algorithm



- Fast Retransmit occurs after three (3) duplicate Acks are received
- Fast Recovery means that CWND is not reduced to IW

Source and Copyright of graph is unknown

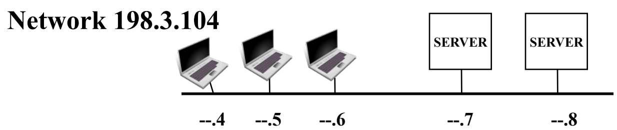
Fast Retransmit via Duplicate ACK



Note that this is a simplified diagram and that TCP Acks and Sequence numbers are based on the byte, not the segment or packet

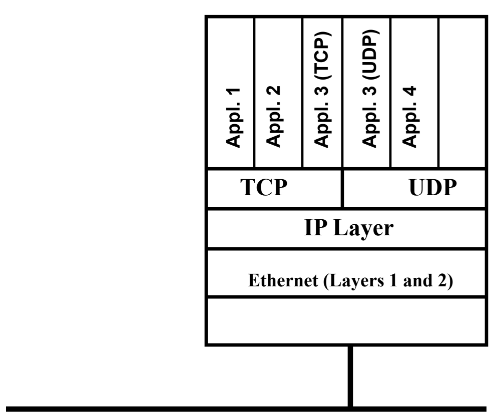
Connection Management

Application Layer Connection Management

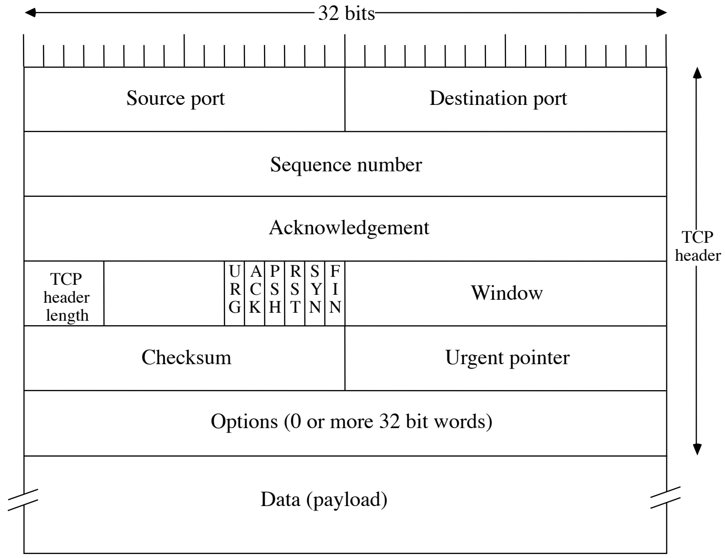


- **How does a system keep track of all of its application layer connections?**
- **Can we see the details of these connections?**

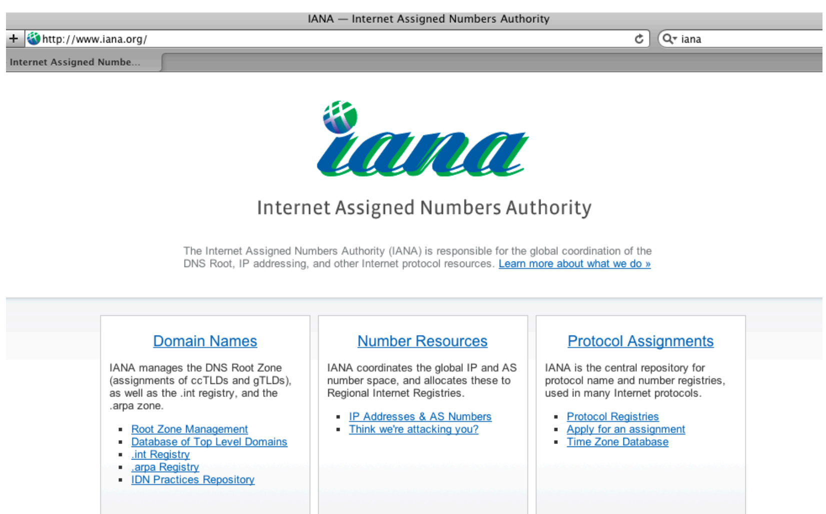
Application Layer Software Schematic



TCP Packet Header



WWW.IANA.ORG

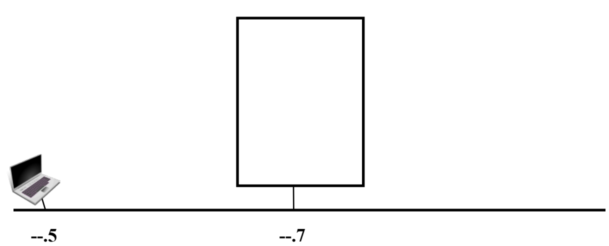


Some Well Known TCP Port Numbers

20,21	FTP	File transfer
22	SSH	Secure Shell
23	Telnet	Remote login, not encrypted
25	SMTP	Email
80	HTTP	world wide web
110	POP3	Remote email access
443	HTTPS	Encrypted web traffic
1720	H.323	Video conferencing
5060	SIP	Session Initiation Protocol

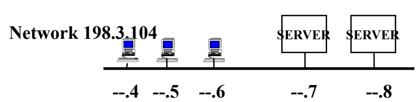
(SIP for VoIP also uses multiple dynamic ports)

Application Layer Connection Management (2)



Network 198.3.104

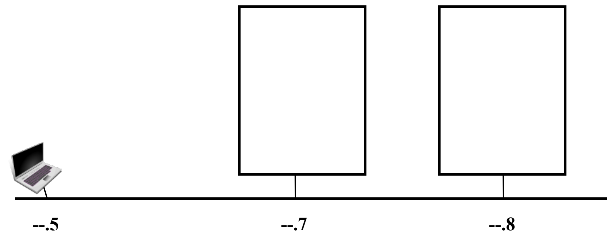
Connection Management Table



Connection ID #					

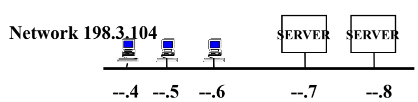
© 1998 - 2017 L. Evenchik

Application Layer Connection Management (2)



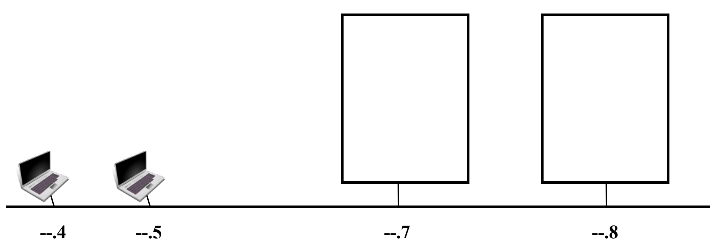
Network 198.3.104

Connection Management Table



Connection ID #	Protocol (TCP/UDP)	Local IP	Remote IP	Local Port	Remote Port

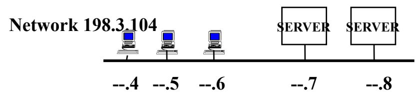
Application Layer Connection Management (2)



Network 198.3.104

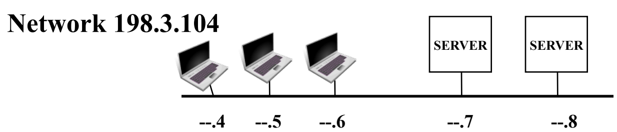
© 1998 - 2017 L. Evenchik

Connection Management Table



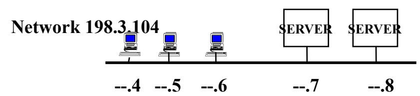
Connection ID #	Protocol (TCP/UDP)	Local IP	Remote IP	Local Port	Remote Port

Application Layer Connection Management



- **How does a system keep track of all of its application layer connections?**
- **Can we see the details of these connections?**

Connection Management Table



Connection ID #	Protocol (TCP/UDP)	Local IP	Remote IP	Local Port	Remote Port

netstat -an

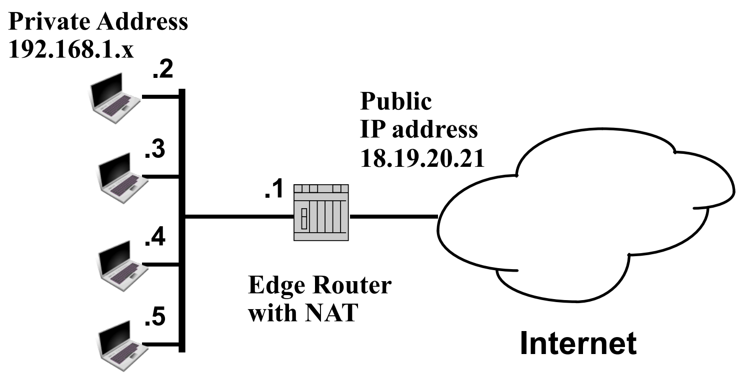
tcp	140.247.30.107.80	24.60.123.123.1518	ESTABLISHED
tcp	140.247.30.107.23	24.60.234.234.2055	ESTABLISHED
tcp	140.247.30.107.25	24.60.222.221.2006	ESTABLISHED
tcp	140.247.30.107.110	134.174.111.222.1186	FIN_WAIT_2
tcp	140.247.30.107.143	134.174.123.213.1682	ESTABLISHED
tcp	140.247.30.107.80	134.174.212.121.1683	ESTABLISHED
tcp	140.247.30.107.22	24.60.33.22.1516	TIME_WAIT
tcp	*.80	*.*	LISTEN
tcp	*.443	*.*	LISTEN
tcp	*.22	*.*	LISTEN
tcp…..			

Network Address Translation (NAT)

Network Address Translation (NAT)

- NAT functionality maps between private addresses and public addresses using various mechanisms. Remember that private means a non-routable address within the Internet.
- NAT breaks the Internet's end-to-end model.
- NAT functionality can be standalone or implemented in routers, proxies, application layer gateways (ALG), firewalls, SBC, etc.
- Operational details vary by the type of protocol (ICMP, UDP, TCP) as well as type of application layer protocol (email, HTTP, FTP, peer-to-peer, SIP, H.323, mapping, voice and video)
- How do you create and manage the table in the NAT that keeps track of connections? This is what we need to understand.

Network Address Translation (NAT) Basic Topology Diagram



Routers, Firewalls, and other network devices implement NAT today.

Private IP Addresses

The common private network address prefixes

• 10/8	10.0.0.0 to 10.255.255.255
• 172.16/12	172.16.0.0 to 172.31.255.255
• 169.254/16	169.254.0.0 to 169.254.255.255 Used for Local Link Address
• 192.168/16	192.168.0.0 to 192.168.255.255

RFC 6890 - Special-Use IPv4 Addresses

IANA IPv4 Special-Purpose Addre...

IANA IPv4 Special-Purpose Address Registry

Created
2009-08-19
Last Updated

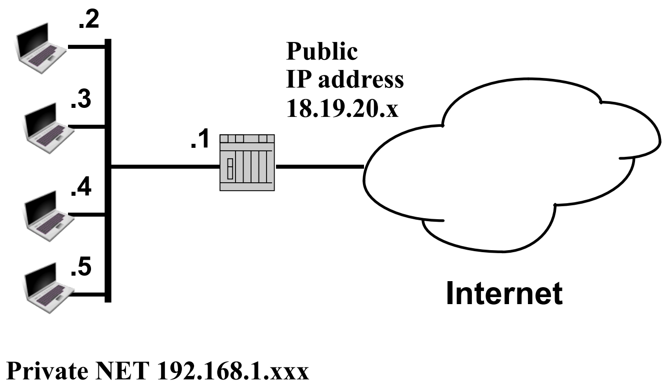
Address Block	Name	RFC	Allocation Date	Termination Date	Source	Destination	Forwardable	Global	Reserved-by-Protocol
0.0.0.0/8	"This host on this network"	[RFC1122], section 3.2.1.3	September 1981	N/A	True	False	False	False	True
10.0.0.0/8	Private-Use	[RFC1918]	February 1996	N/A	True	True	True	False	False
100.64.0.0/10	Shared Address Space	[RFC6598]	April 2012	N/A	True	True	True	False	False
127.0.0.0/8	Loopback	[RFC1122], section 3.2.1.3	September 1981	N/A	False[1]	False[1]	False[1]	False[1]	True
169.254.0.0/16	Link Local	[RFC3927]	May 2005	N/A	True	True	False	False	True
172.16.0.0/12	Private-Use	[RFC1918]	February 1996	N/A	True	True	True	False	False
192.0.0.0/24[2]	IETF Protocol Assignments	[RFC6890], section 2.1	January 2010	N/A	False	False	False	False	False
192.0.0.0/29	DS-Lite	[RFC6333]	June 2011	N/A	True	True	True	False	False
192.0.0.170/32, 192.0.0.171/32	NAT64/DNS64 Discovery	[RFC-ietf-behave-nat64-discovery-heuristic-17], section 2.2	February 2013	N/A	False	False	False	False	True
192.0.2.0/24	Documentation (TEST-NET-1)	[RFC5737]	January 2010	N/A	False	False	False	False	False
192.88.99.0/24	6to4 Relay Anycast	[RFC3068]	June 2001	N/A	True	True	True	True	False
192.168.0.0/16	Private-Use	[RFC1918]	February 1996	N/A	True	True	True	False	False
198.18.0.0/15	Benchmarking	[RFC2544]	March 1999	N/A	True	True	True	False	False
198.51.100.0/24	Documentation (TEST-NET-2)	[RFC5737]	January 2010	N/A	False	False	False	False	False
203.0.113.0/24	Documentation (TEST-NET-3)	[RFC5737]	January 2010	N/A	False	False	False	False	False
240.0.0.0/4	Reserved	[RFC1112], section 4	August 1989	N/A	False	False	False	False	True
255.255.255.255/32	Limited Broadcast	[RFC919], section 7	October 1984	N/A	False	True	False	False	False

Also review RFC 6598 (April 2012) called
IANA-Reserved IPv4 Prefix for Shared Address Space

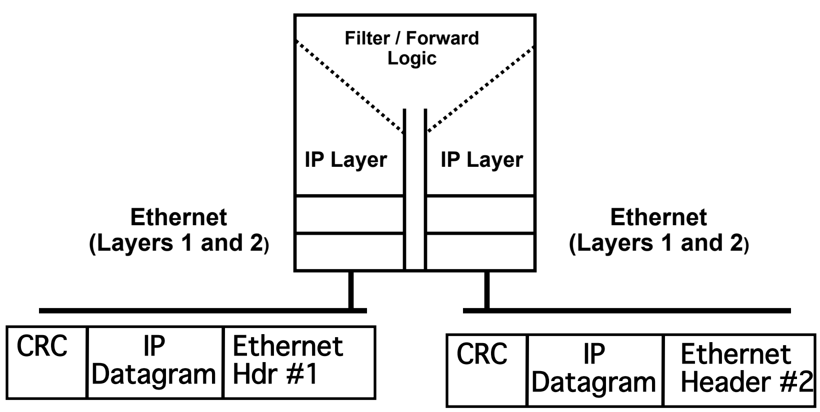
Network Address Translation (NAT)

- 1-to-1 address mapping
- Many-to-many address mapping
- Network Address and Port mapping, called NAPT or PAT

Network Address Translation (NAT) Block Diagram

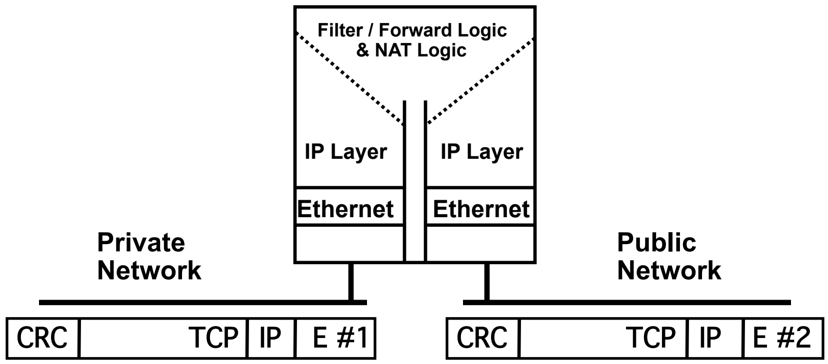


Remember, Ethernet Headers are Built Separately on Each Side of a Router



Simple NAT Functionality Implemented by a Router

(Packet sent from private to public network.)



IP Source Addr – Private address
IP Destination Addr – Public Dest.

IP Source Addr – Mapped to Public address
IP Destination Addr – Public Dest. (no change)

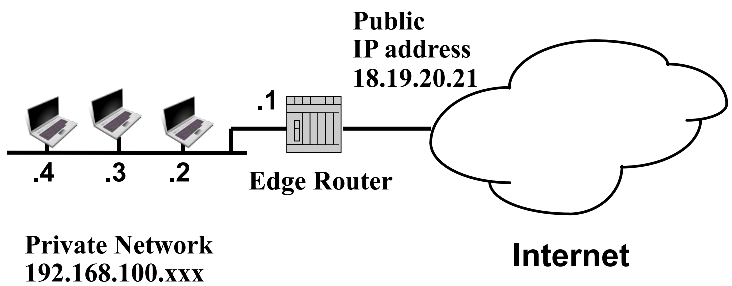
Simplified Connection ID Table

Connection ID #	Local IP Address	Local Port	Foreign IP Address	Foreign Port	Protocol

Simplified NAT Table in Router/Firewall

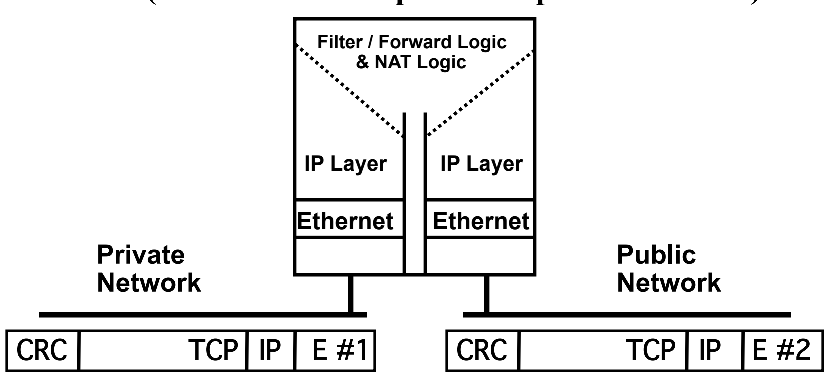
Connection ID #					

Network Address Port Translation (NAPT) Block Diagram



Simple NAT Functionality Implemented by a Router

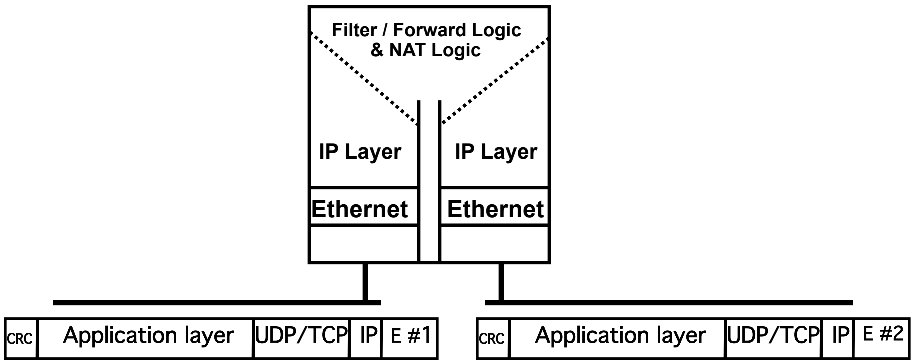
(Packet sent from private to public network.)



IP Source Addr – Private address
IP Destination Addr – Public Dest.

IP Source Addr – Mapped to Public address
IP Destination Addr – Public Dest. (no change)

NAT and Port-Mapping Functionality Implemented by a Router or Firewall



IP Source Addr – Private
IP Dest Addr – Public
TCP Source Port – Private
TCP Dest Port - Public

IP Source Addr – Mapped to Public address
IP Destination Addr – Public
TCP Source Port – Mapped to Public port
TCP Dest Port - Public

Simplified NAPT/PAT Table

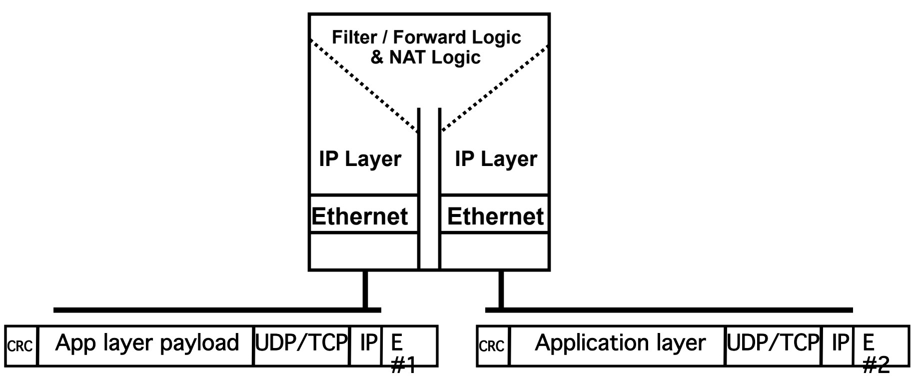
Simplified NAPT/PAT Table in Firewall

Protocol	Private Source IP Address	Private Source Port	Mapped Source IP Address	Mapped Source Port	Foreign IP Address	Foreign Port #

NAT - Outstanding Issues (1 of 2)

- NAT breaks the Internet's end-to-end connection model. How and why this happened and its importance is still discussed and debated. The same questions are now being asked about IPv6.
- As we will learn later in the term there is a significant problem with embedded addresses at the application layer as used in SIP and VoIP. Note though that this is not a new problem since it was an issue with FTP, which is over 40 years old.)
- There are issues with IPsec and other end-to-end security mechanisms.
- There is a problem of managing incoming connections of any type as well as long term UDP connections. Lets think about this.

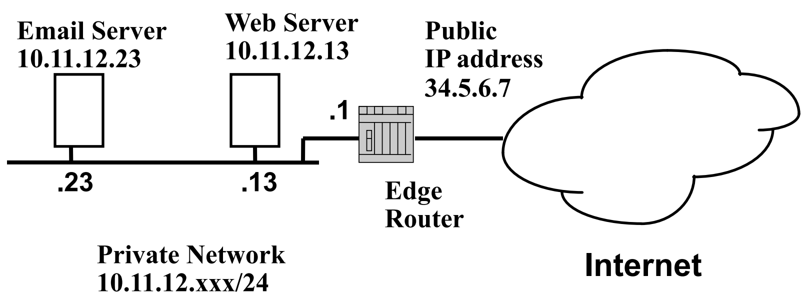
NAT and Port-Mapping with Embedded Addresses in Application Layer



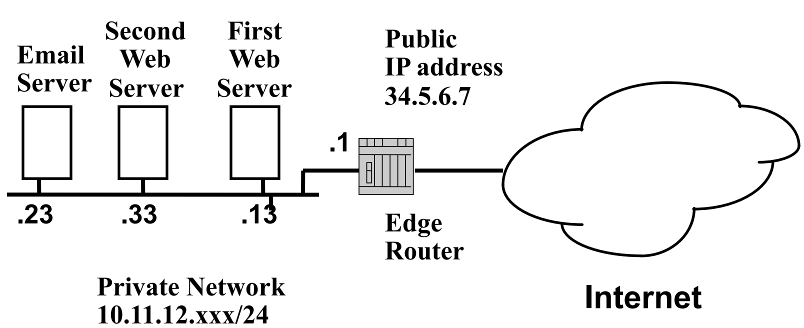
IP Source Addr – Private
IP Dest Addr – Public
TCP Source Port – Private
TCP Dest Port – Public
App Layer – application layer can also contain embedded addresses (SIP, H.323)

IP Source Addr – Mapped (Public)
IP Destination Addr – Public
TCP Source Port – Mapped (Public)
TCP Dest Port – Public
App Layer – application layer can also contain embedded addresses (SIP, H.323)

Handling Incoming Connections with NAT



Handling Incoming Connections with NAT (Multiple Web Servers in DMZ)



NAT - Outstanding Issues (2 of 2)

- Network management and debugging is much more difficult with NAT (as you would expect.)
- NAT is NOT a solution to network security! It just provides some address obscurity.
- Regardless, NAT is as common as the home router.
- NAT is also being used as part of the transition to IPv6.

Some Questions related to NAT

- What behavior do your users see if the edge router crashes? How does it depend on the protocol, HTTP versus SSH for example?
- What happens if UDP or TCP is not used, such as with ICMP?
- What about protocols such as VoIP (SIP) that open multiple connections, or require an inbound connection?
- How should we support inbound connections to systems such as a web server? What happens if you want to support more than one web server?
- What happens when there are two levels of NAT? This is now very common with wireless networks.

Simplified NAPT/PAT Table

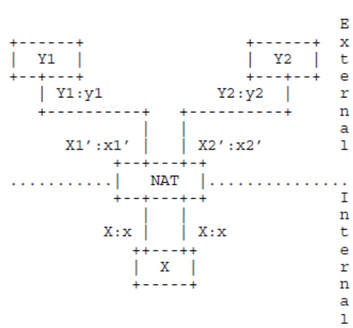
Protocol	Private Source IP Address	Private Source Port	Mapped Source IP Address	Mapped Source Port	Foreign IP Address	Foreign Port #

NAT Related Protocols: RFCs 3022, 4787, 5382, 5389

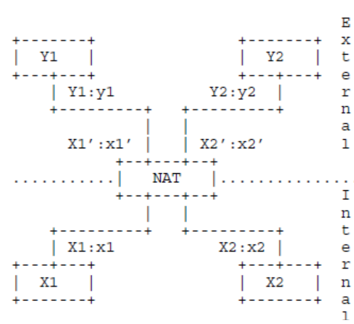
RFCs about NAT

- RFCs 4787 and 5382 further define the type of filtering, address mapping and port mapping that is done by a NAT device for UDP and TCP. They also specify behaviors.
- It is obviously not as simple or as consistent as the description that we have started with in this lecture
- Address and port mapping behaviors:
  – Endpoint-Independent Mapping
  – Address-Dependent Mapping
  – Address and Port-Dependent Mapping
- Filtering behaviors:
  – Endpoint-Independent Filtering
  – Address-Dependent Filtering
  – Address and Port-Dependent Filtering

NAT Reference Topologies, RFC 4787



Address and Port Mapping



Port Assignment

STUN, TURN and ICE

STUN, RFC 5389

- One of a number of approaches and protocols designed to allow a client to determine if and how NAT is being used. Vendors have had proprietary approaches for this for many years and there are websites that allow users to determine their IP address (such as whatismyip)
- STUN - Session Traversal Utilities for NAT. The original RFCs called it Simple Traversal of UDP Through Network Address Translators, but it did not provide the hoped for functionality so it was significantly revised.
- STUN is a lightweight request/response protocol that can discover the presence and type of NATs in the network
- Requires both client software and STUN servers on the external network
- Protocol does not try to solve the problem of incoming connections through NAT.
- STUN is intended to work in concert with TURN and ICE
- nat-stun-port is 3478 for both UDP and TCP

STUN Architecture

- STUN client software embedded in a device that needs to know about the presence of NAT. A good example is a VoIP phone connected to a corporate or home network.
- STUN server is typically located in the carrier (ISP or Telco) network or the public Internet
- STUN servers have multiple IP addresses
- Multiple STUN servers should be available for reliability
- STUN servers can be discovered via DNS SRV records (RFC 2782)

STUN Protocol Operation

(non-secure operation)

- To determine the presence of NAT:
  - STUN client sends a request to STUN server
  - STUN server copies (and encodes) source IP address and port from IP and UDP headers into the STUN payload and sends response back to client
  - Client compares its own IP address and port number to returned information in the payload
- To determine the type of NAT:
  - Client sends another request to the STUN server's second IP address and then compares returned (NAT) IP and port info to the first request
  - Client asks server to respond using second IP address and different port to determine NAT operation

Quick Definition of STUN, TURN and ICE

- STUN – protocol used by a client to determine the presence and type of NAT
- TURN – protocol for working with a media relay which is typically located on the public Internet. (We'll cover this later when we talk about SIP and VoIP.)
- ICE – protocol and technique for dealing with and managing NAT traversal in protocols such as SIP (for VoIP) that use the offer/answer model. We'll cover this later when we talk about SIP and VoIP.

One Minute Wrap-Up

- Please do this Wrap-Up at the end of each lecture.
- Please fill out the form on the website.
- The form is anonymous (but you can include your name if you want.)
- Please answer three questions:
  - What is your grand "Aha" for today's class?
  - What concept did you find most confusing in today's class?
  - What questions should I address next time
- **Thank you!**

© 1998 - 2017 L. Evenchik