Internet Traffic Monitoring and Analysis using NG-MON

Seminar at HP Labs, Palo Alto, CA, USA
Dec. 4, 2003

James Won-Ki Hong

Distributed Processing & Network Management Lab.

Dept. of Computer Science and Engineering

POSTECH, Korea

jwkhong@postech.ac.kr
http://dpnm.postech.ac.kr/~jwkhong
Tel: +82-54-279-2244

* This work was in part supported by HY-SDR Research Center at Hanyang University under the ITRC program of Ministry of Information and Communication, Korea.

Table of Contents

1. Introduction

2. NG-MON: Next Generation Network Traffic MONitoring and Analysis System

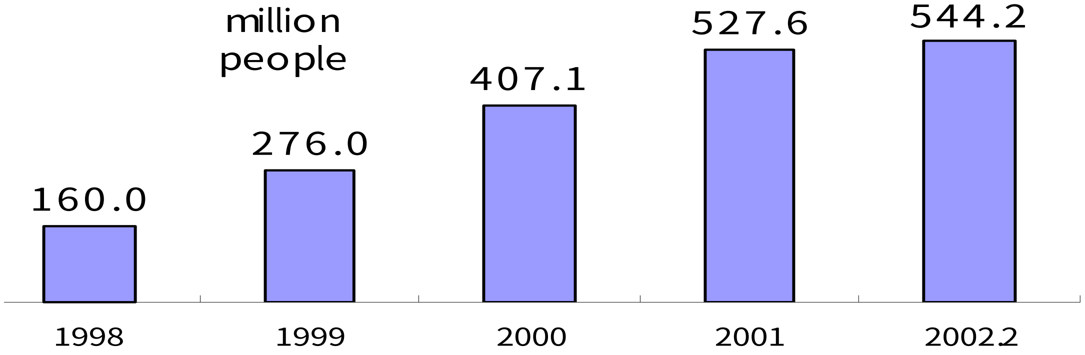
3. Traffic Analysis Methods
   - Multimedia Service Traffic
   - Peer-to-Peer (P2P) Traffic

4. Implementation, Deployment & Usage

5. Summary

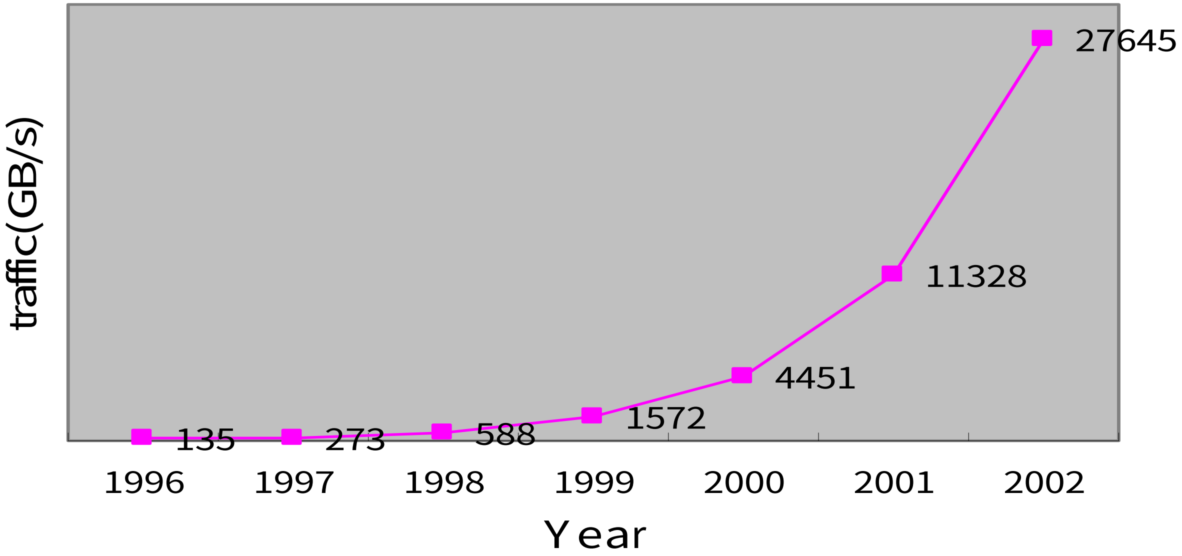
1. Introduction – Growth of Internet Use

The number of Internet users is growing



Source : Nua Inc.

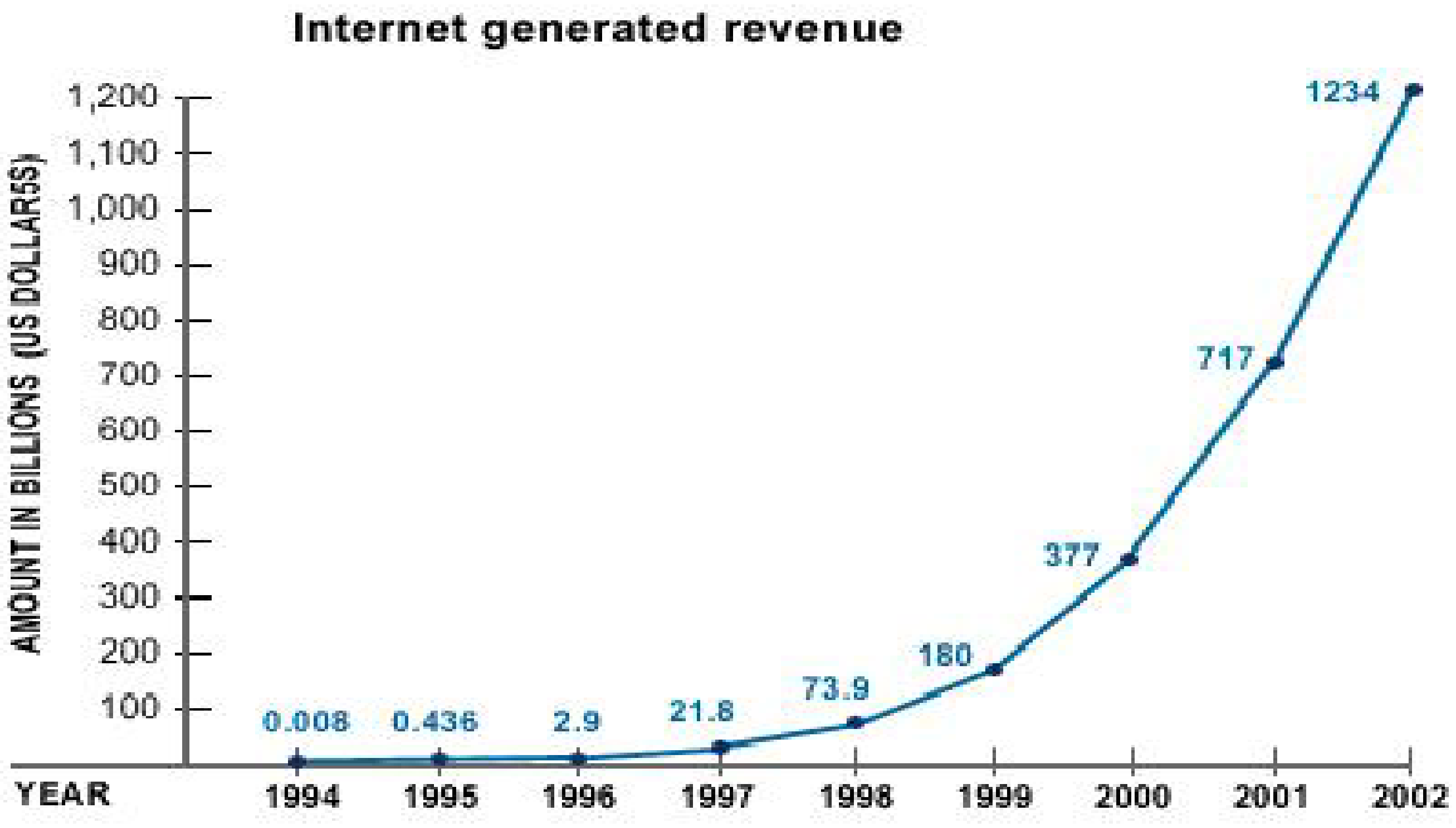
Internet traffic has increased dramatically



Source: America's Network

1. Introduction – Reliance on Internet

The Internet generated revenue has been increasing rapidly!



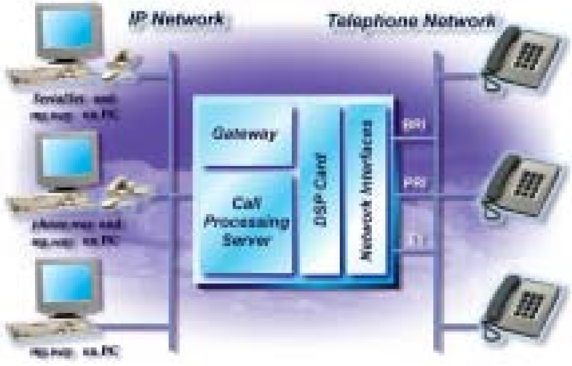
Source : Active Media.

❖ Internet's importance and reliance are increasing!

1. Introduction – Internet Applications

- ❖ Traditional Internet Applications
  - ➢ Web, FTP, Email, Telnet, etc.
- ❖ Emerging Internet applications
  - ➢ Online games, shopping, banking, stock trading, network storage
  - ➢ VOD, EOD, VoIP
  - ➢ P2P applications – instant messaging, file sharing

Online game

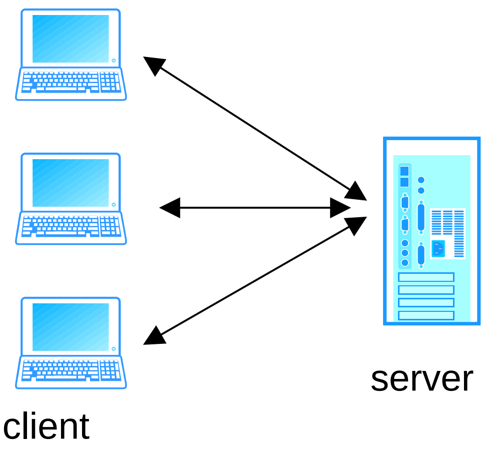


VoIP

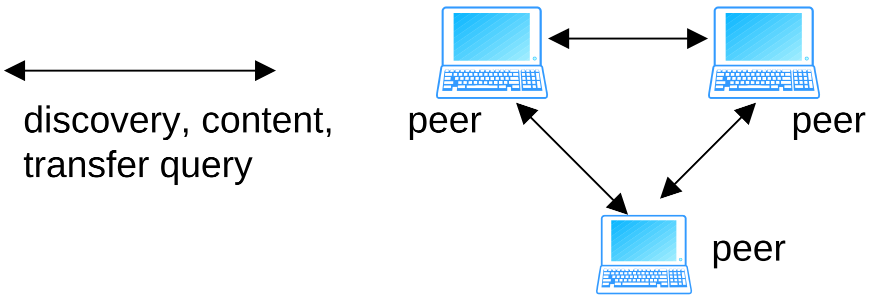
VOD

1. Introduction – Structure of Applications

- Client-Server
  - Traditional structure



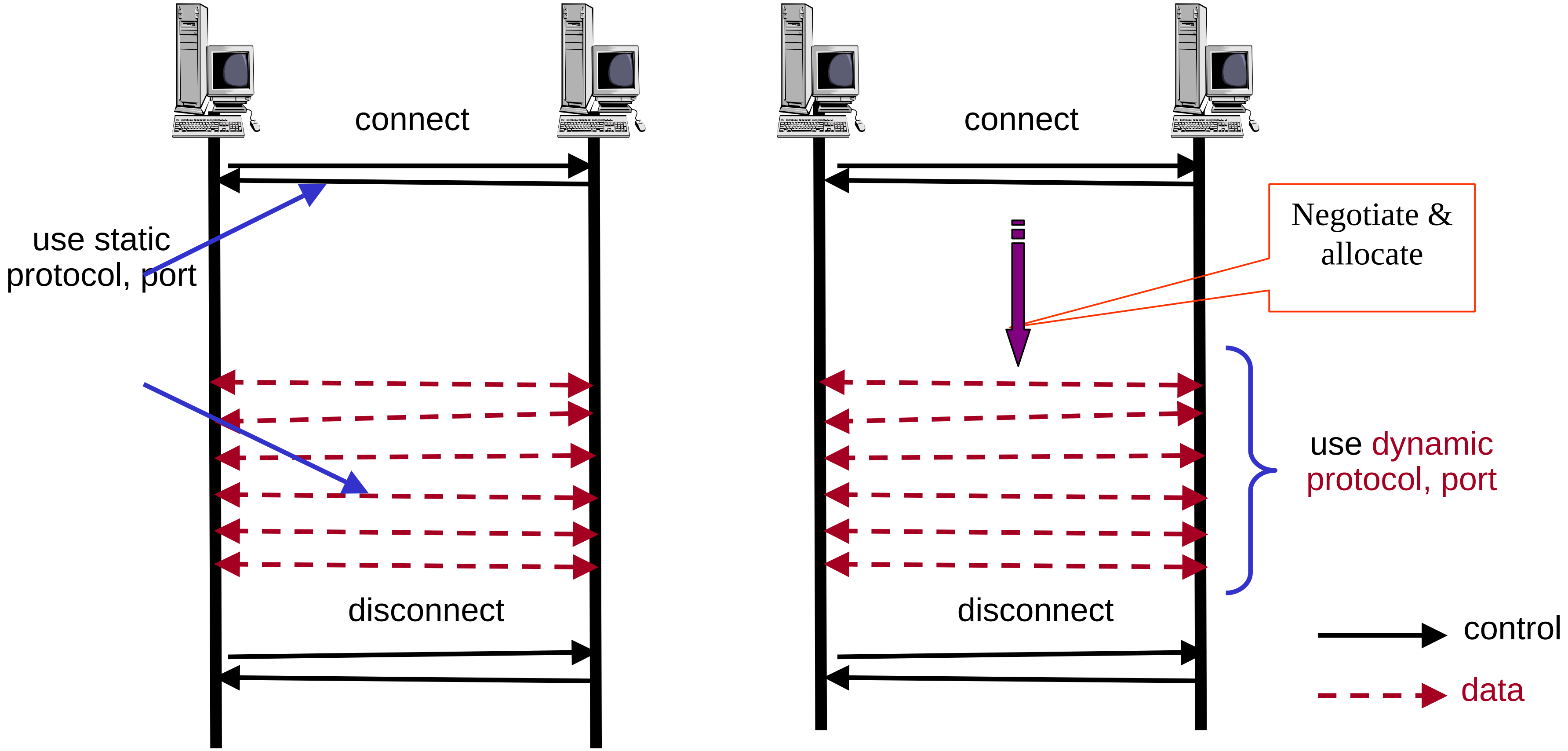
- Peer-to-Peer (P2P)
  - New concept for messaging and file sharing
  - Generates high volume of traffic



- Structures of applications are changing!

1. Introduction – Types of Traffic

- Static sessions vs. Dynamic sessions



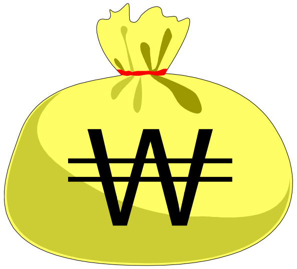
1. Introduction – Motivation

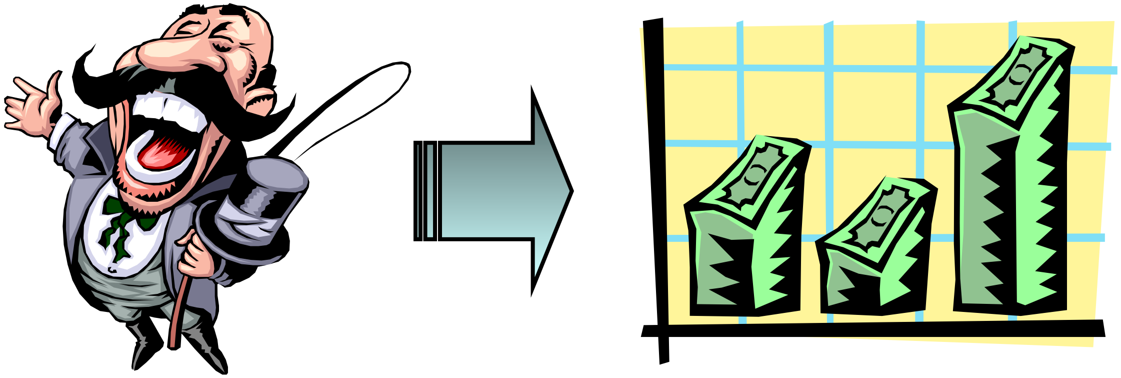
- Needs of Users
  - Want to get their money's worth
  - **Fast, reliable, high-quality, secure, virus-free** Internet access

- Needs of Service Providers
  - Understand the behavior of their networks
  - Provide fast, high-quality, reliable service to satisfy customers and thus **reduce churn rate**
  - Plan for network deployment and expansion
  - SLA monitoring
  - Network security attack detection and prevention

1. Introduction – Application Areas

- Network Problem Determination and Analysis
- Traffic Report Generation
- Intrusion & Hacking Attack (e.g., DoS, DDoS) Detection
- Service Level Monitoring (SLM)
- Network Planning
- Usage-based Billing
- Customer Relationship Management (CRM)





1. Introduction – Research Problems

❖ Capturing Packets

- ➢ How to capture all packets from high-speed, high volume networks (Mbps→Gbps→Tbps)?

❖ Flow Generation & Storage

- ➢ What packet info to save to perform various analysis?
- ➢ How to minimize storage requirements?

❖ Analysis

- ➢ How to analyze and generate information needed quickly?
- ➢ Streaming media (Windows Media, Real, Quicktime)
- ➢ Multimedia Conferencing, VoIP
- ➢ P2P & game traffic
- ➢ Network Security Attacks (Internet Worms & Viruses)

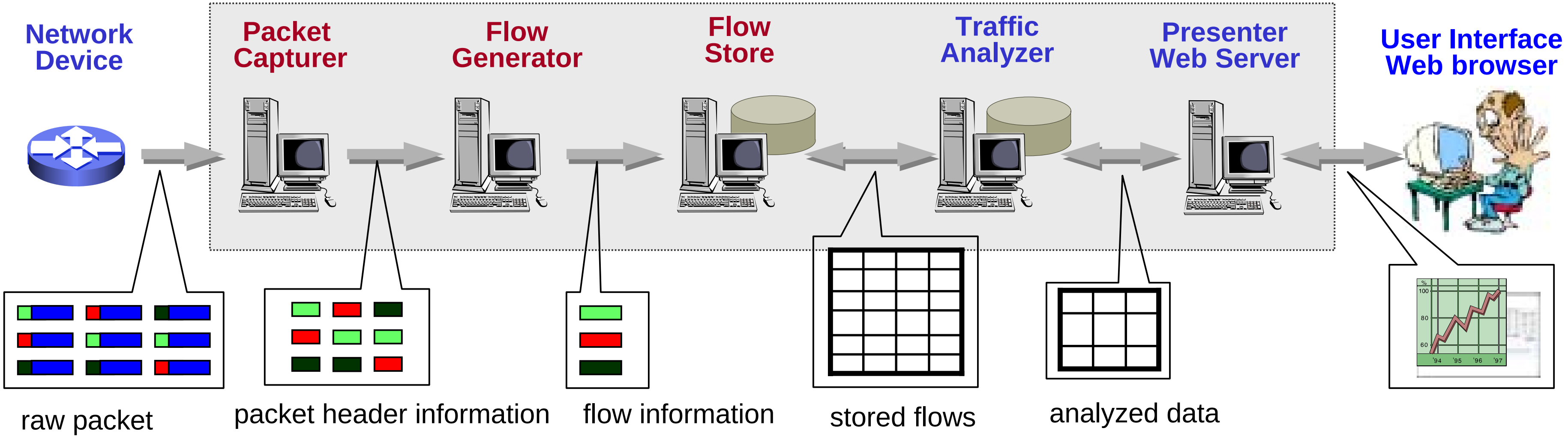
2. NG-MON

- Our previous work
  - MRTG+ (1996-97)
    - Traffic load analysis with sensitive map
  - WebTrafMon-I (1997-98)
    - Traffic type analysis on a single monolithic system (up to 10 Mbps)
  - WebTrafMon-II (1999-2001)
    - Traffic type analysis using a distributed architecture (up to 100 Mbps)
- NG-MON (2002-present)
  - **N**ext **G**eneration Network Traffic **MON**itoring and Analysis System
  - Targeting 10 Gbps or higher networks
  - To support various analysis applications
    - Streaming media, multimedia conferencing, P2P, game traffic analysis
    - Network security attack detection and analysis
    - SLA monitoring
    - Usage-based billing
    - Customer relationship management

NG-MON - Requirements

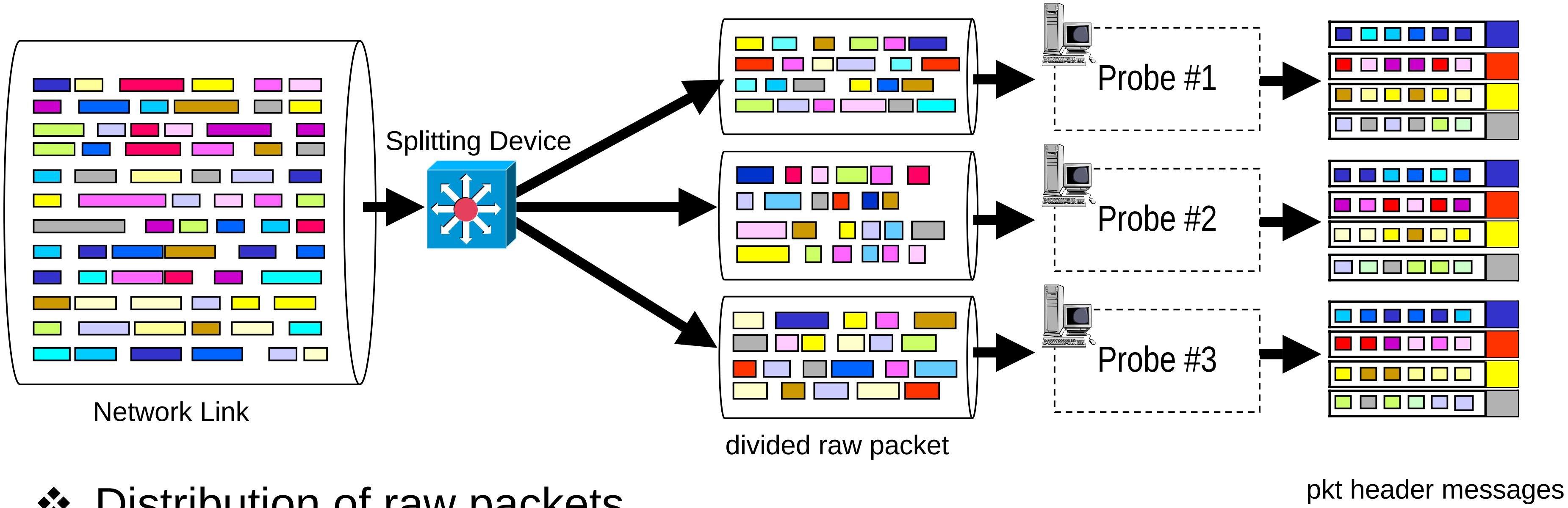
- Distributed, load-balancing architecture for scalability
  - subdivide monitoring system into several functional components
  - efficient load sharing between phases and within each phase
  - pipelined and parallel architecture
- Lossless packet capture
- Flow-based analysis
  - aggregate packet information into flows for efficient processing
- Considerations for small storage requirements
- Support for various applications

NG-MON - Design



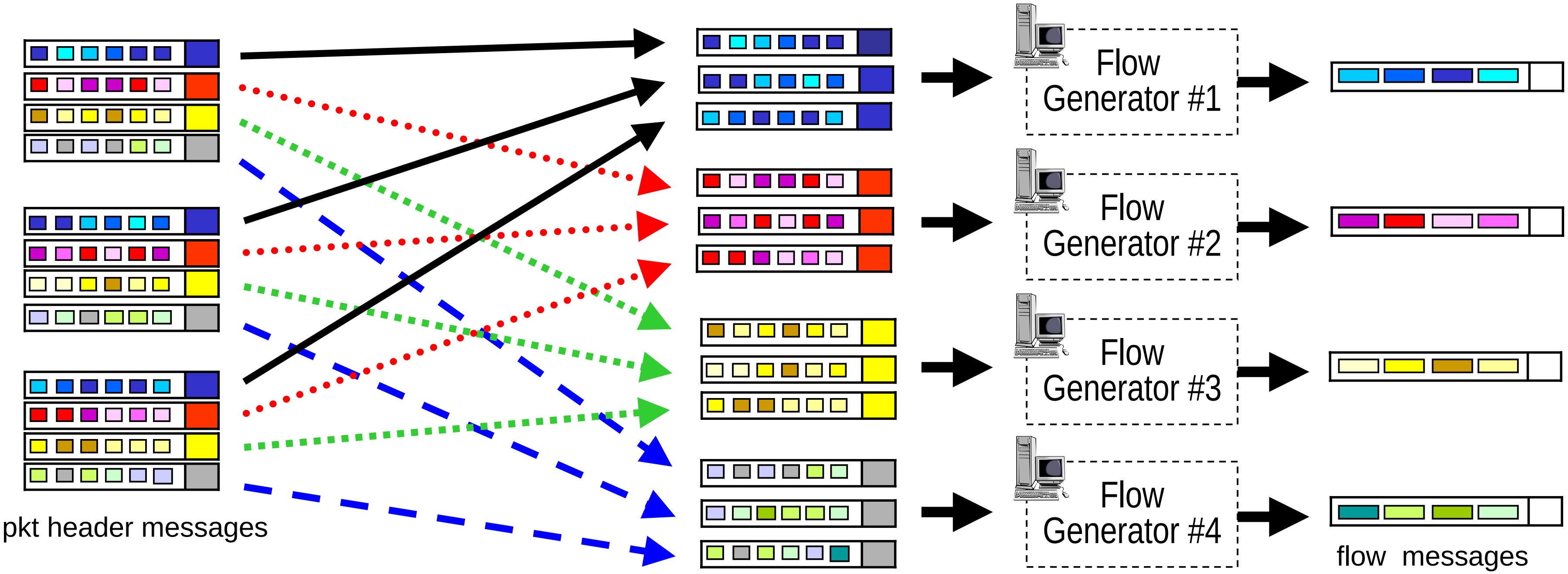
- ❖ NG-MON is composed of 5 phases
  - ➢ Packet Capture
  - ➢ Flow Generation
  - ➢ Flow Store
  - ➢ Traffic Analysis
  - ➢ Presentation & Reporting

NG-MON - Packet Capture



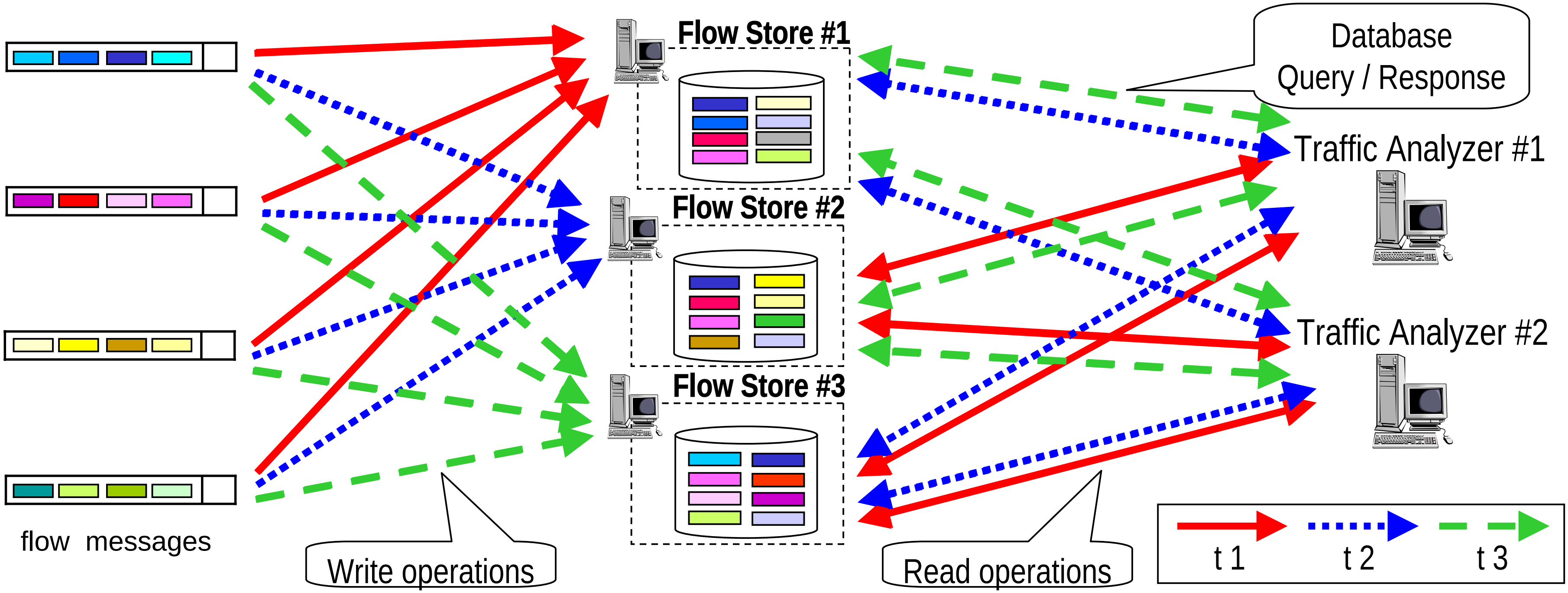
- ❖ Distribution of raw packets
  - ➢ by using splitting function provided by an optical splitter
  - ➢ by using mirroring function provided in network devices
- ❖ Probe
  - ➢ captures all packets coming into probe
  - ➢ export buffer-queues: one to one with flow generators
  - ➢ fills buffer-queues with packet header’s 5-tuple based hashing
  - ➢ collect the scattered packets in the same flow into the same buffer-queue

NG-MON - Flow Generation



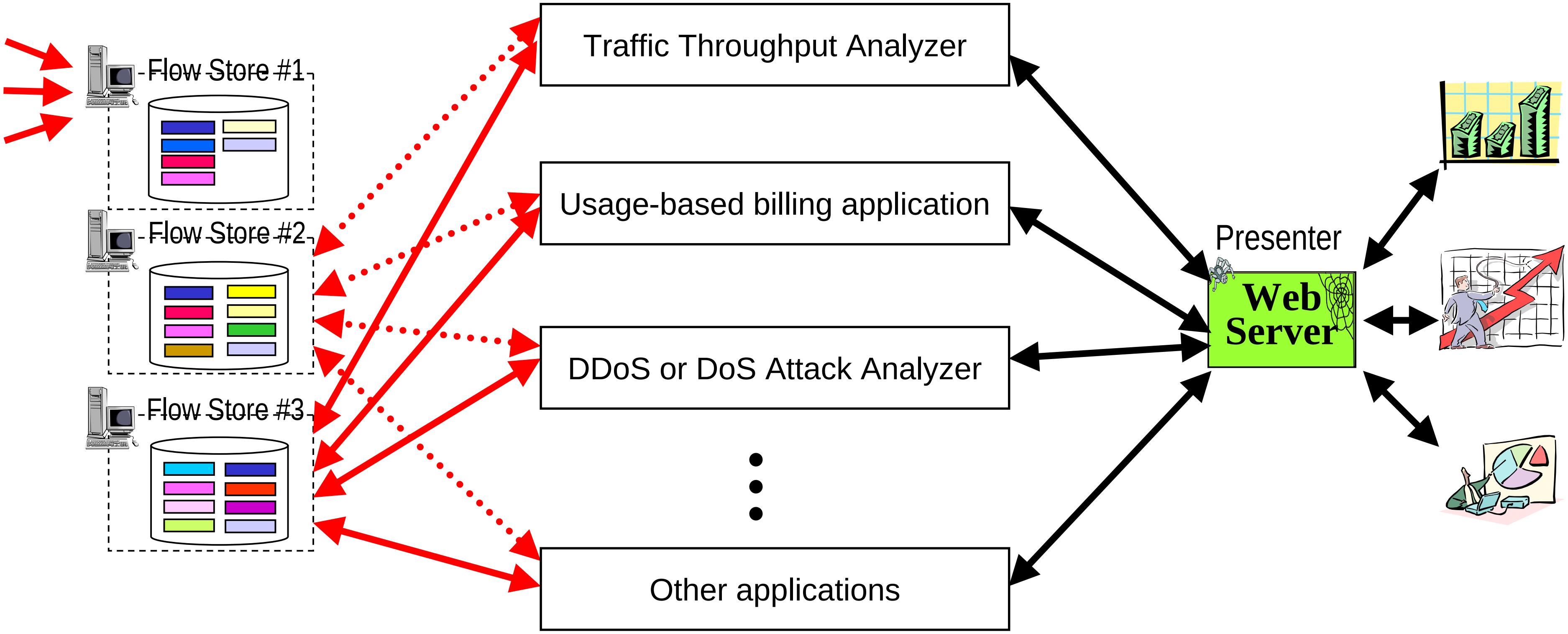
- Distribution of packet header information
  - 5-tuple based hashing in the probe
  - Packet header messages of potentially the same flow get delivered to the same flow generator
- Flow generator receives packet header messages and generates flows and exports flow messages to flow store

NG-MON - Flow Store



- Separation of write operations from read operations
  - the destination address of flow message is assigned to the flow store according to the time
  - While one or more flow stores are inserting flow data, the other flow stores are queried by the traffic analyzers
- Flow store provides traffic information to support various analysis applications
  - provides an analysis API to analyzers

NG-MON - Traffic Analysis & Presentation



Flow Store #1
Flow Store #2
Flow Store #3
Traffic Throughput Analyzer
Usage-based billing application
DDoS or DoS Attack Analyzer
Other applications
Presenter
Web Server

- Analyzer extracts information from Flow Stores and can perform application specific analysis
- Separate analyzer is needed for each application

NG-MON - Implementation

Phase	Packet Capture	Flow Generator	Flow Store	Analyzer	Presenter
Development Tool	pcap library C language	C language	C language MySQL	C language MySQL	PHP jpgraph library
Hardware System	▪ Xeon 2.4 GHz 2 CPUs ▪ 1 Gbytes memory ▪ 2-1000 Mbps NICs ▪ 80 GB hard disk	▪ Pentium-III 800 MHz CPU ▪ 256 Mbytes memory ▪ 2-100 Mbps NICs ▪ 20GB hard disk			
OS	Redhat Linux 7.2				

3. Application Traffic Analysis - Problems

- Newly emerging various Internet applications
  - Streaming media applications
  - Game applications
  - P2P applications
- New structures of Internet applications
- Various application level protocols
  - Not standard, not publicly open
- Use of Dynamic Ports
- Use of Multiple sessions

Streaming Media Traffic Analysis (1/3)

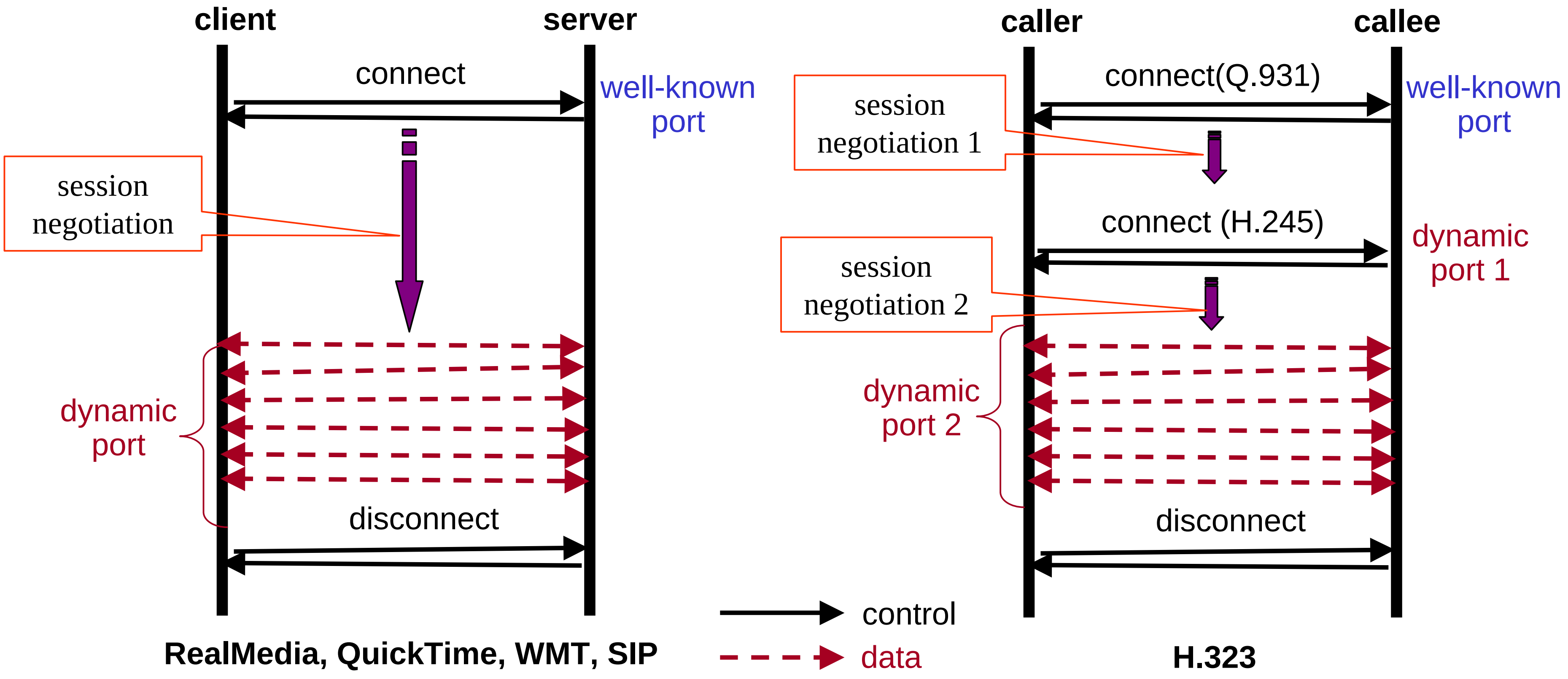
- Services and Protocols
  - **Control protocol**: setup/close connection, fast forward/backward
  - **Data transfer protocol**: transfer multimedia data

Streaming service platform	Control protocol	Data transfer protocol	Service vendor
Real Media	RTSP	RDT	Real Networks
QuickTime	RTSP	RTP	Apple
Windows Media Technology	MMS	MMST or MMSU	Microsoft

Multimedia conferencing	Control protocol	Data transfer protocol	Standard organization
Applications based on H.323	Q.931 H.245	RTP	ITU-T
Applications based on SIP	SIP	RTP	IETF

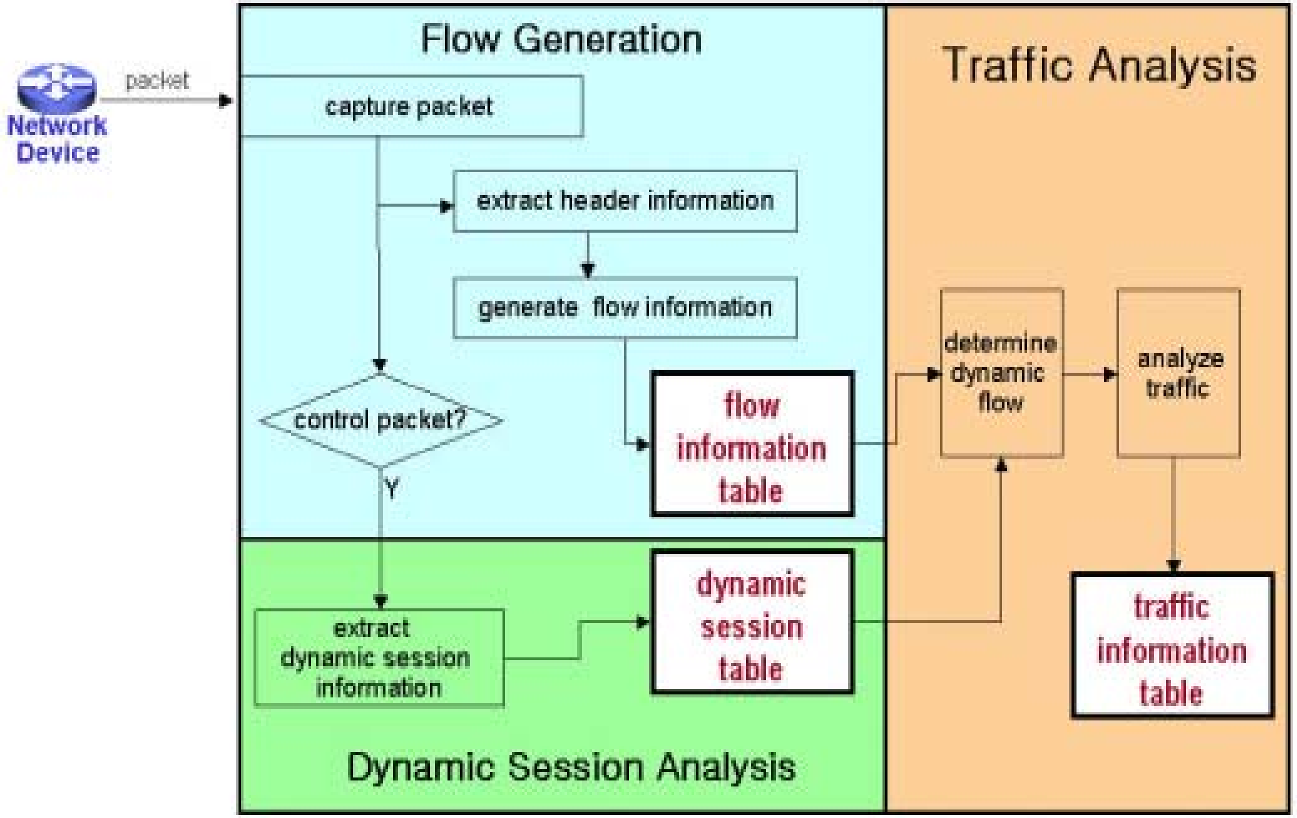
Streaming Media Traffic Analysis (2/3)

- Services and Protocols
  - **Control protocol**: setup/close connection, fast forward/backward
  - **Data transfer protocol**: transfer multimedia data



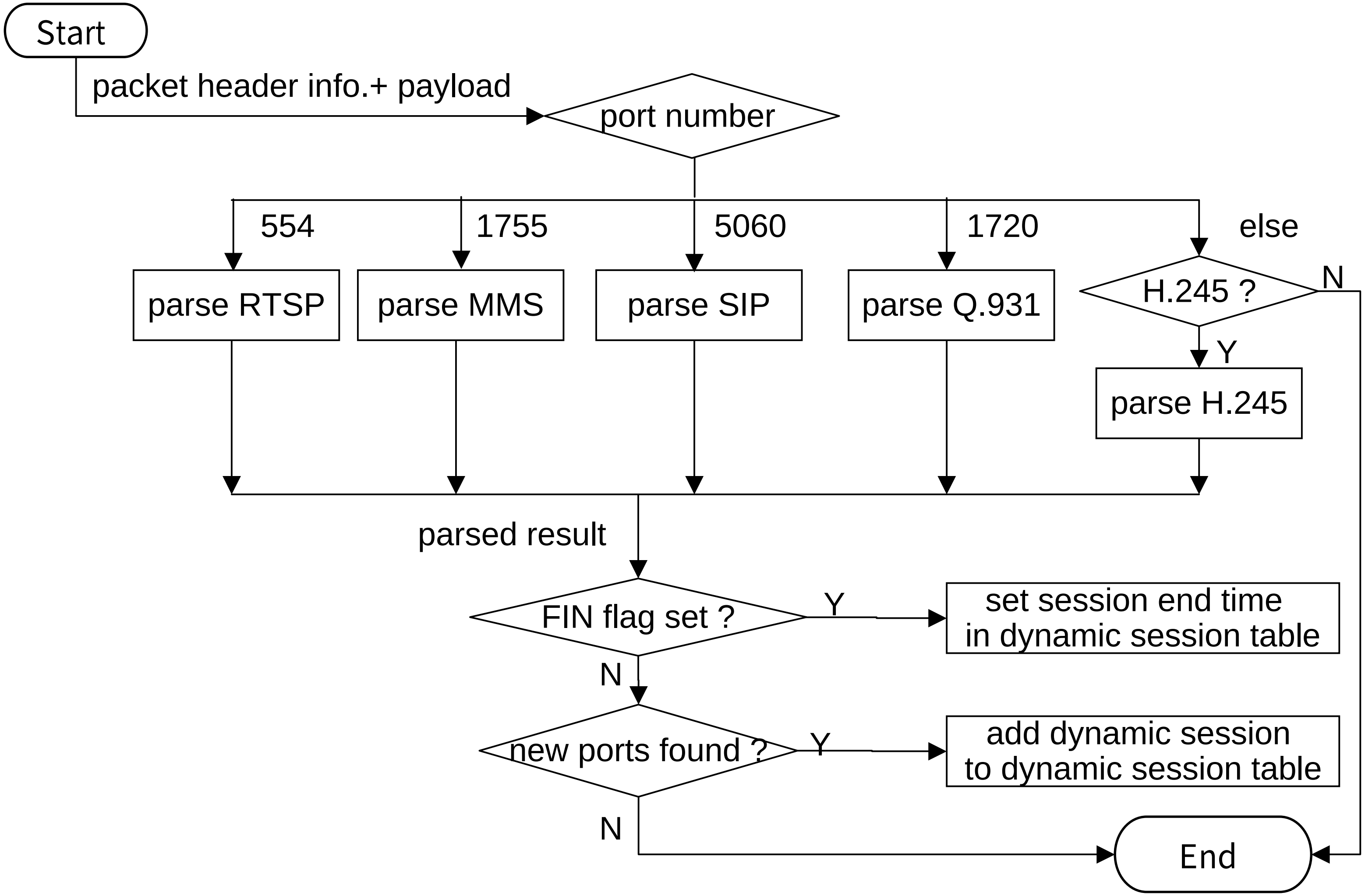
Streaming Traffic Analysis Method (3/3)

- 3 phases of Payload Examination Method
  1. Flow Generation
  2. Dynamic Session Analysis
  3. Multimedia Service Traffic Analysis



Dynamic Session Analysis

- Obtain dynamic session information from control packet

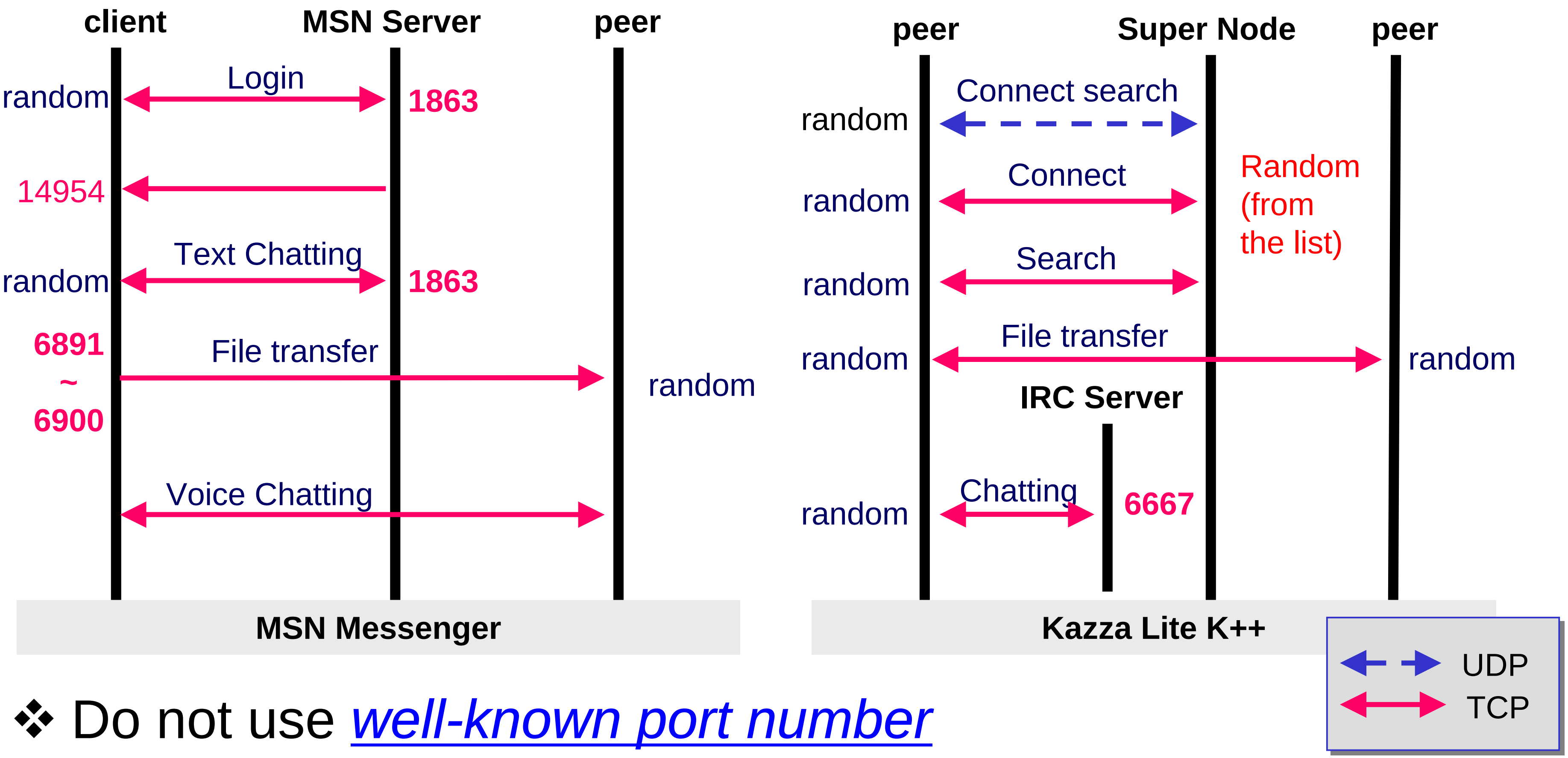


P2P Traffic Analysis (1/3)

- ❖ Two types of P2P applications: Instant messaging & File sharing
- ❖ Large number of P2P applications
- ❖ Various functions supported

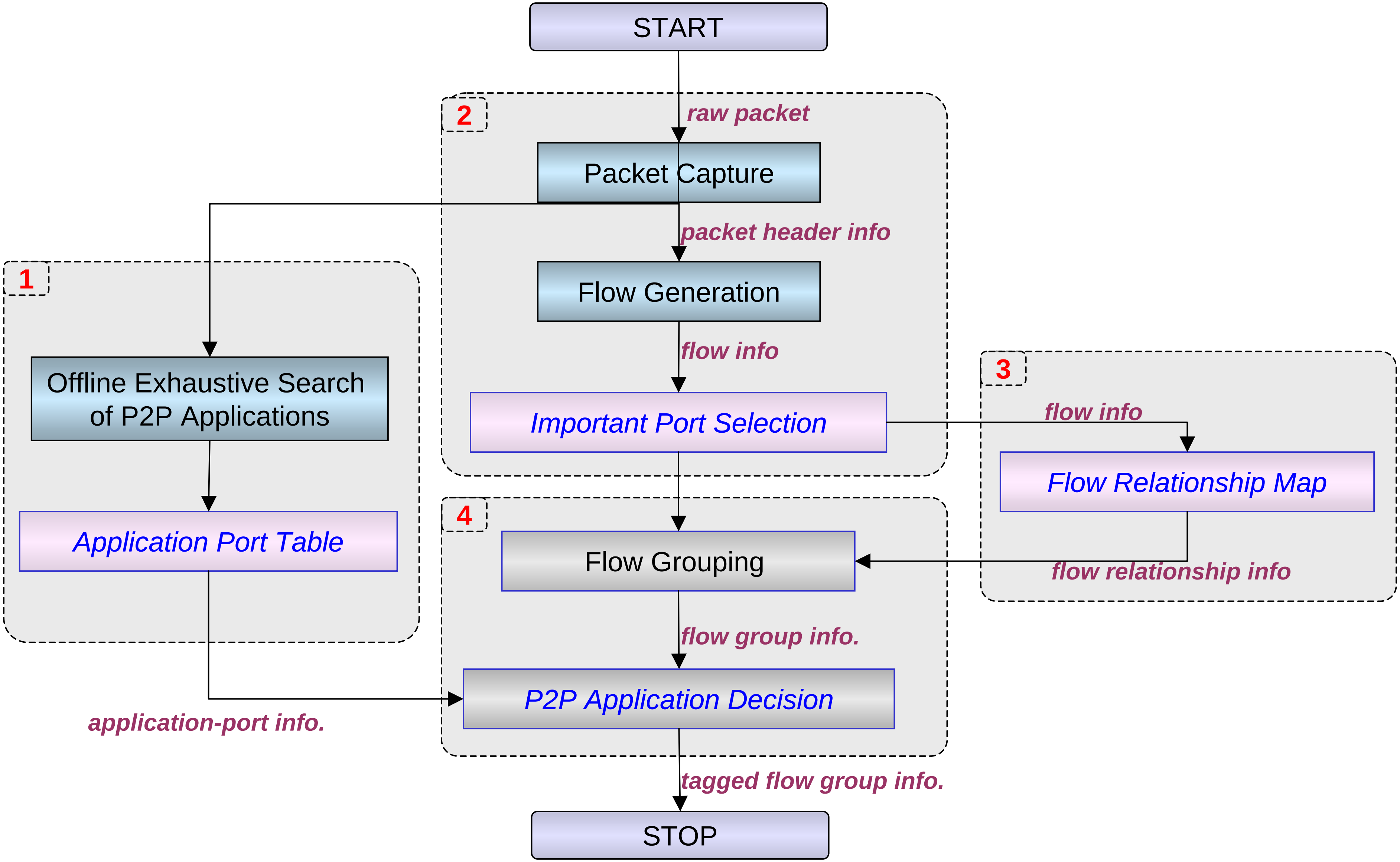
	Instant Messaging Application	**File Sharing Application**
Functions	- Message delivery - 1:1 & multi-chatting - voice & video chatting - File transfer - …	- Searching - File sharing - Chatting - …
Applications	- MSN Messenger - Yahoo Messenger - ICQ, AOL Messenger - Daum Messenger - …	- Kazaa - eDonkey - Gnutella - WinMX - …

P2P Traffic Analysis (2/3)



- ❖ Do not use *well-known port number*
- ❖ Lots of *P2P applications*
- ❖ No standard communication *protocol*
- ❖ Use multiple *sessions* for various functions

P2P Traffic Analysis Method (3/3)



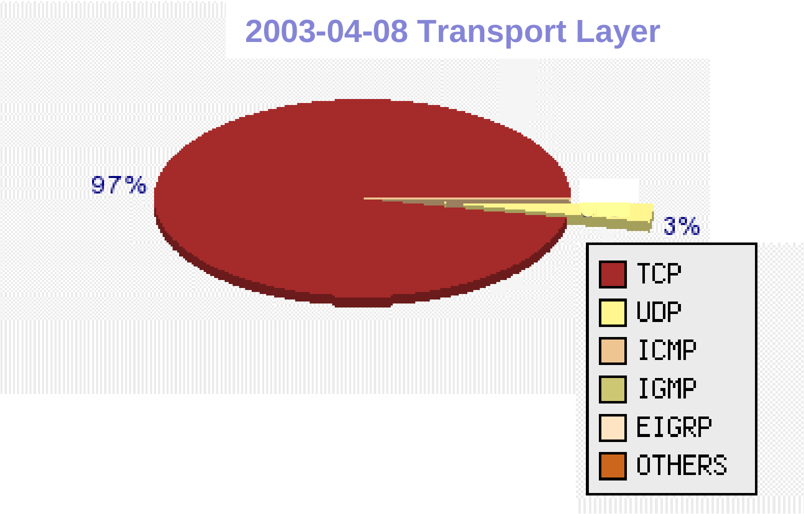
Application Port Table (APT)

- Offline survey of P2P Applications
  - Find out most-frequently used port numbers used by each P2P application
  - Use packet analysis tool like *tcpdump*, *ethereal*
  - Select one port number as a representative port among them

Application Name	TCP		UDP	
	representative port	frequently used ports	representative port	frequently used ports
MSN Messenger	1863	1863, 6981-6990, 14594		
Yahoo Messenger	5101	5101, 5050		
Soribada	22322	22322, 7675, 7676, 7677	22321	22321, 7674
eDonkey	4661	4661, 4662, 6667		
Guruguru	9292	9292, 9999, 31200, 22000, 22400, 21700		
V-share	8404	8403, 8404, 1212, 8903, 8908, 8909, 15561		
Shareshare	6399	6399	6777	6388, 6733, 6777

Important Port Number Selection

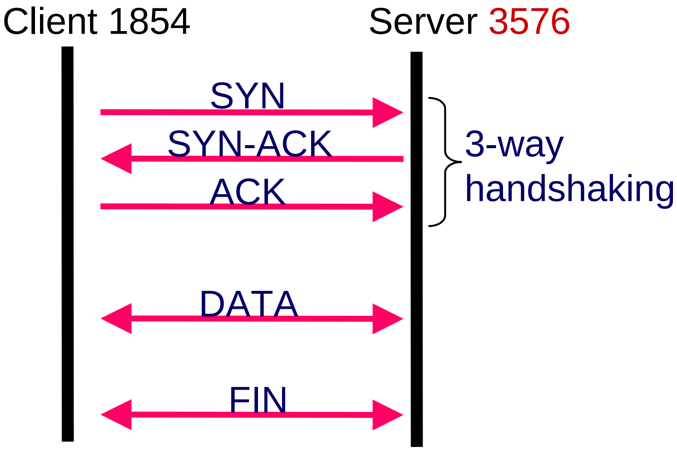
- Most of IP traffic is TCP Traffic
- Most of P2P Traffic is TCP Traffic
- The server listening port is important in the analysis of TCP Traffic



- How to decide server listening port in the captured TCP flow

- Use SYN and SYN-ACK packet
  - SYN packet
    - destination port number
  - SYN-ACK packet
    - source port number

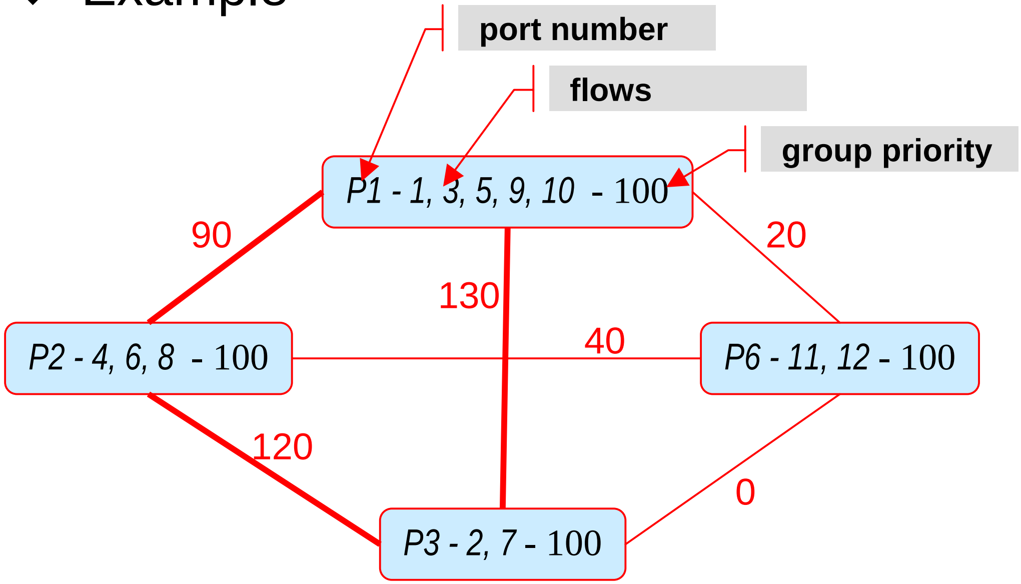
- In case of UDP Port
  - use flow relationship



TCP Communication Sequence

Flow Relationship Map (FRM)

- Property Dependency Grouping (PDG)
  - Use *source port*, *destination port*, and *proto.*
  - Set priority to each combination
- Location Dependency Grouping (LDG)
  - Use *source IP* and *destination IP*
  - Make link among group with priority
- Example



	proto	source port	destination port	priority
0				0
1			1	20
2		1		20
3		1	1	50
4	1			0
5	1		1	50
6	1	1		50
7	1	1	1	100

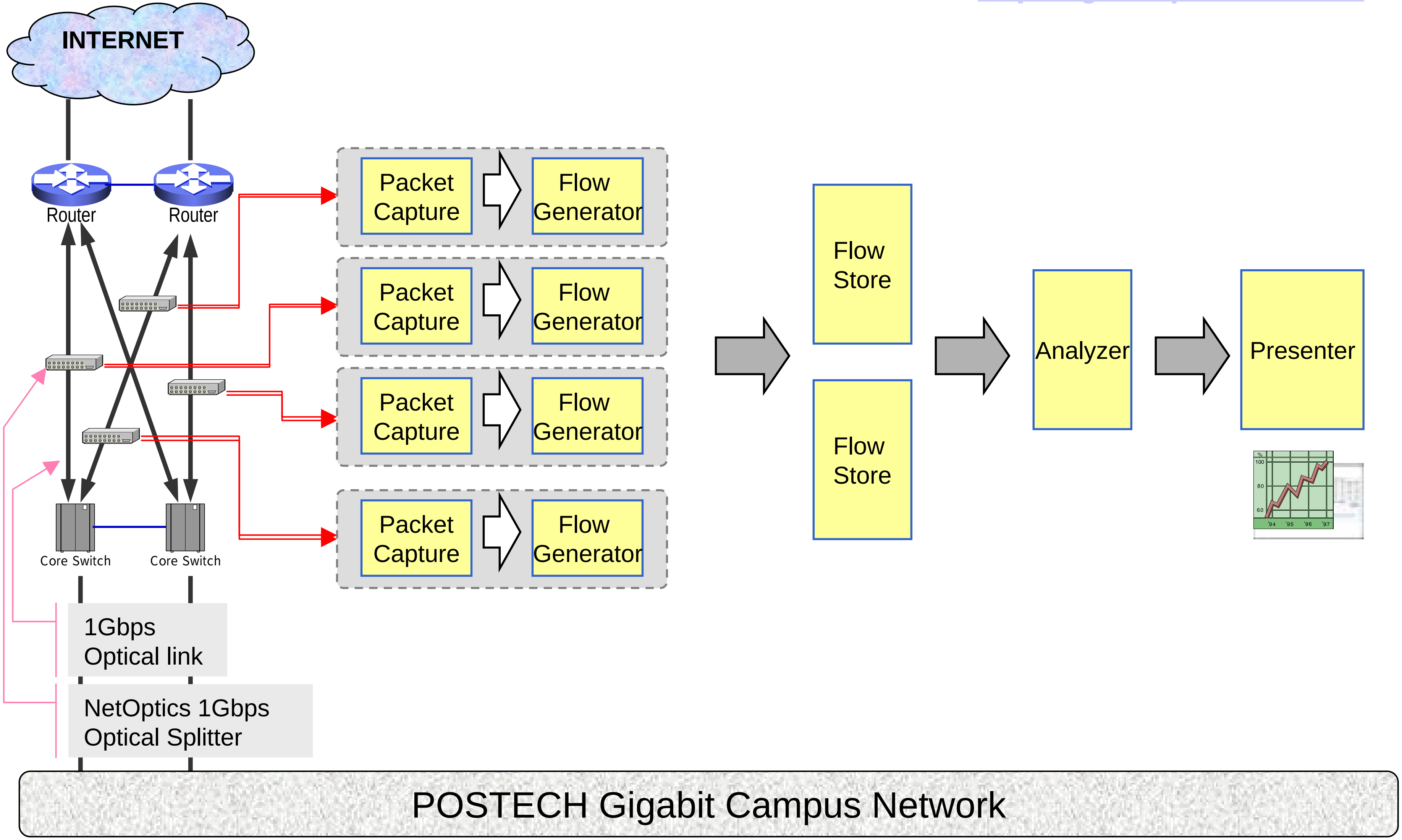
Property Dependency Table

source ip	destination ip	priority
		0
	1	10
1		10
1	1	100

Location Dependency Table

4. NG-MON - Deployment at POSTECH

http://ngmon.postech.ac.kr



INTERNET

Router

Router

Packet Capture

Flow Generator

Packet Capture

Flow Generator

Packet Capture

Flow Generator

Packet Capture

Flow Generator

Flow Store

Flow Store

Analyzer

Presenter

Core Switch

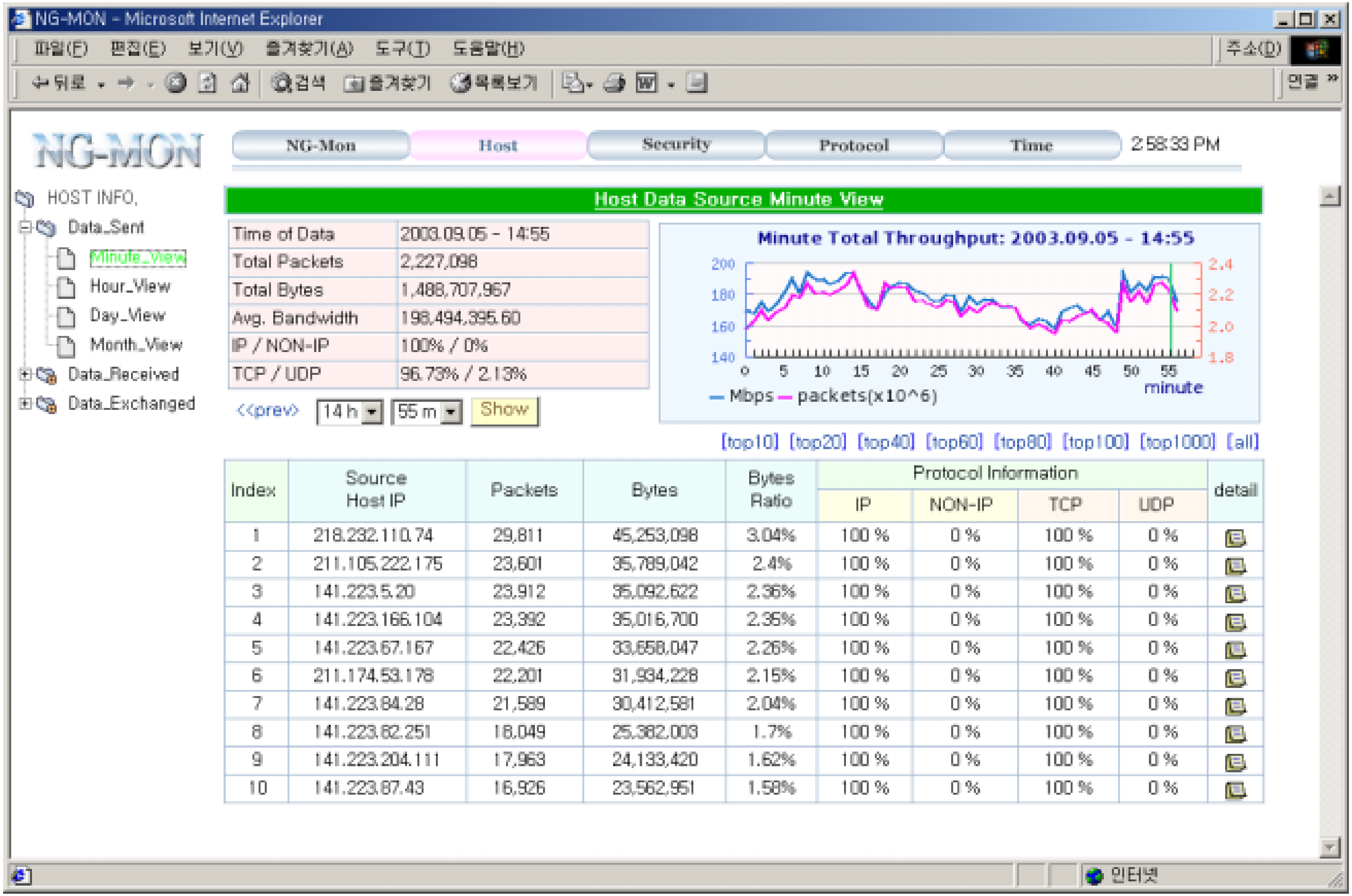
Core Switch

1Gbps Optical link

NetOptics 1Gbps Optical Splitter

POSTECH Gigabit Campus Network

NG-MON - Host Data Sent Minute View



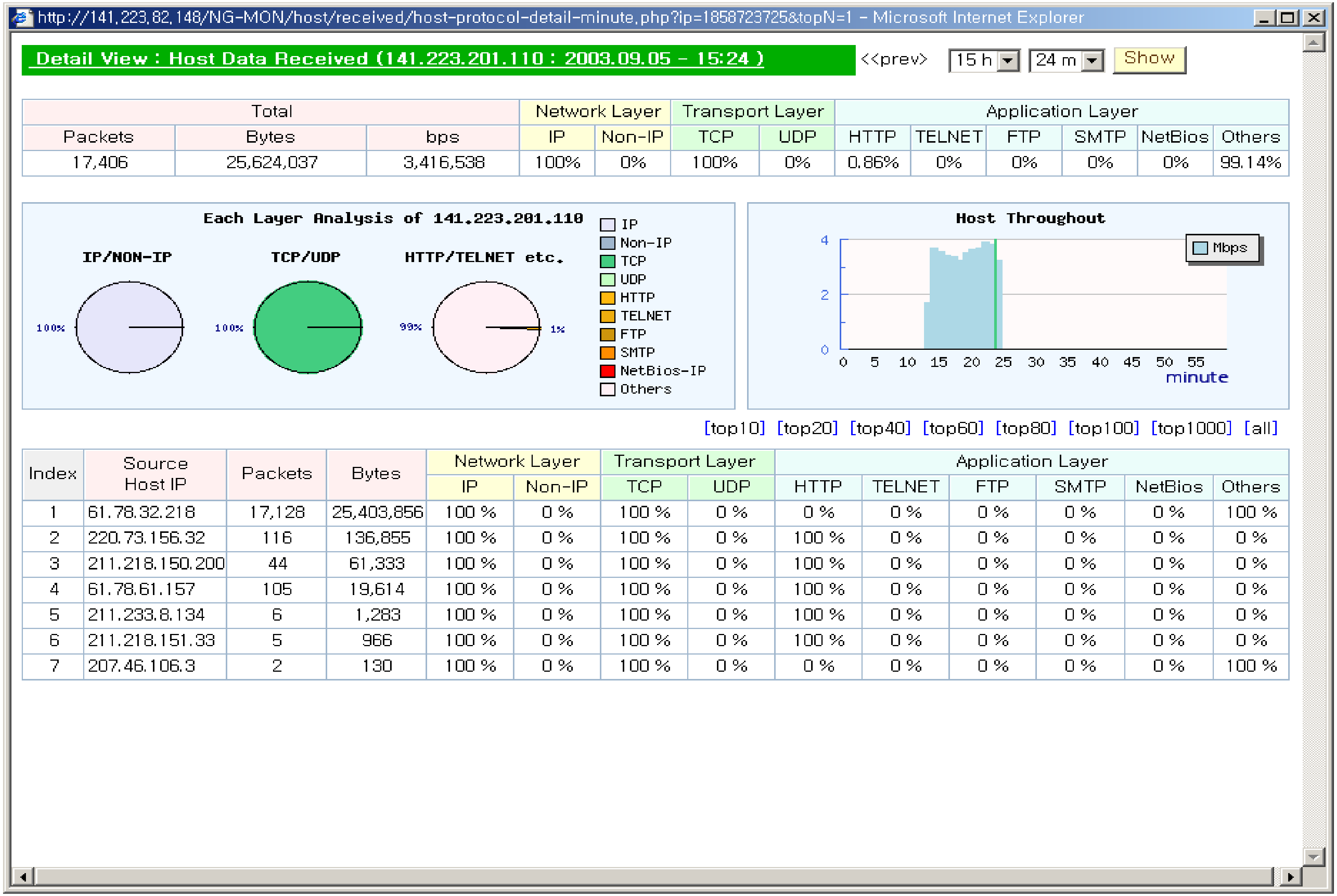
Host Data Source Minute View

Time of Data	2003.09.05 - 14:55
Total Packets	2,227,098
Total Bytes	1,488,707,967
Avg. Bandwidth	198,494,395.60
IP / NON-IP	100% / 0%
TCP / UDP	96.73% / 2.13%

[top10] [top20] [top40] [top60] [top80] [top100] [top1000] [all]

Index	Source Host IP	Packets	Bytes	Bytes Ratio	Protocol Information				detail
					IP	NON-IP	TCP	UDP	
1	218.232.110.74	29,811	45,253,098	3.04%	100 %	0 %	100 %	0 %	
2	211.105.222.175	23,601	35,789,042	2.4%	100 %	0 %	100 %	0 %	
3	141.223.5.20	23,912	35,092,622	2.36%	100 %	0 %	100 %	0 %	
4	141.223.166.104	23,392	35,016,700	2.35%	100 %	0 %	100 %	0 %	
5	141.223.67.167	22,426	33,658,047	2.26%	100 %	0 %	100 %	0 %	
6	211.174.53.178	22,201	31,934,228	2.15%	100 %	0 %	100 %	0 %	
7	141.223.84.28	21,589	30,412,581	2.04%	100 %	0 %	100 %	0 %	
8	141.223.82.251	18,049	25,382,003	1.7%	100 %	0 %	100 %	0 %	
9	141.223.204.111	17,963	24,133,420	1.62%	100 %	0 %	100 %	0 %	
10	141.223.87.43	16,926	23,562,951	1.58%	100 %	0 %	100 %	0 %	

NG-MON - Detailed Host Data Received Minute View



http://141.223.82.148/NG-MON/host/received/host-protocol-detail-minute.php?ip=1858723725&topN=1 - Microsoft Internet Explorer

Detail View : Host Data Received (141.223.201.110 : 2003.09.05 - 15:24)

<<prev> 15 h 24 m Show

Total			Network Layer		Transport Layer		Application Layer					
Packets	Bytes	bps	IP	Non-IP	TCP	UDP	HTTP	TELNET	FTP	SMTP	NetBios	Others
17,406	25,624,037	3,416,538	100%	0%	100%	0%	0.86%	0%	0%	0%	0%	99.14%

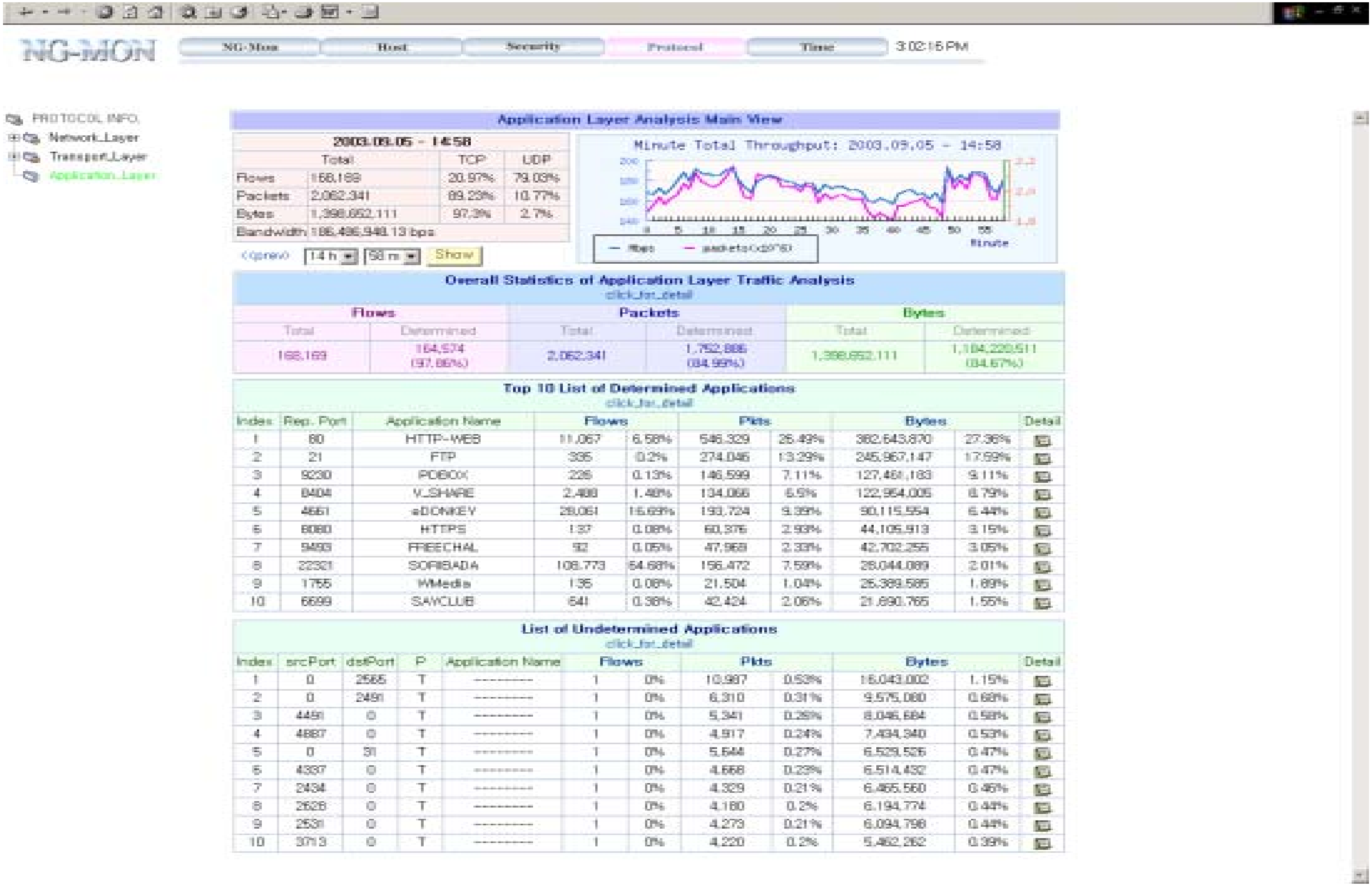
Each Layer Analysis of 141.223.201.110

Host Throughout

[top10] [top20] [top40] [top60] [top80] [top100] [top1000] [all]

Index	Source Host IP	Packets	Bytes	Network Layer		Transport Layer		Application Layer					
				IP	Non-IP	TCP	UDP	HTTP	TELNET	FTP	SMTP	NetBios	Others
1	61.78.32.218	17,128	25,403,856	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	100 %
2	220.73.156.32	116	136,855	100 %	0 %	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
3	211.218.150.200	44	61,333	100 %	0 %	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
4	61.78.61.157	105	19,614	100 %	0 %	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
5	211.233.8.134	6	1,283	100 %	0 %	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
6	211.218.151.33	5	966	100 %	0 %	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %
7	207.46.106.3	2	130	100 %	0 %	100 %	0 %	0 %	0 %	0 %	0 %	0 %	100 %

NG-MON - Application Protocol Minute View



Application Layer Analysis Main View

2003.09.05 - 14:58	Total	TCP	UDP
Flows	168,169	20.97%	79.03%
Packets	2,062,341	89.23%	10.77%
Bytes	1,398,652,111	97.3%	2.7%
Bandwidth	186,486,948.13 bps		

Overall Statistics of Application Layer Traffic Analysis
click_for_detail

Flows		Packets		Bytes	
Total	Determined	Total	Determined	Total	Determined
168,169	164,574 (97.86%)	2,062,341	1,752,886 (84.99%)	1,398,652,111	1,184,220,511 (84.67%)

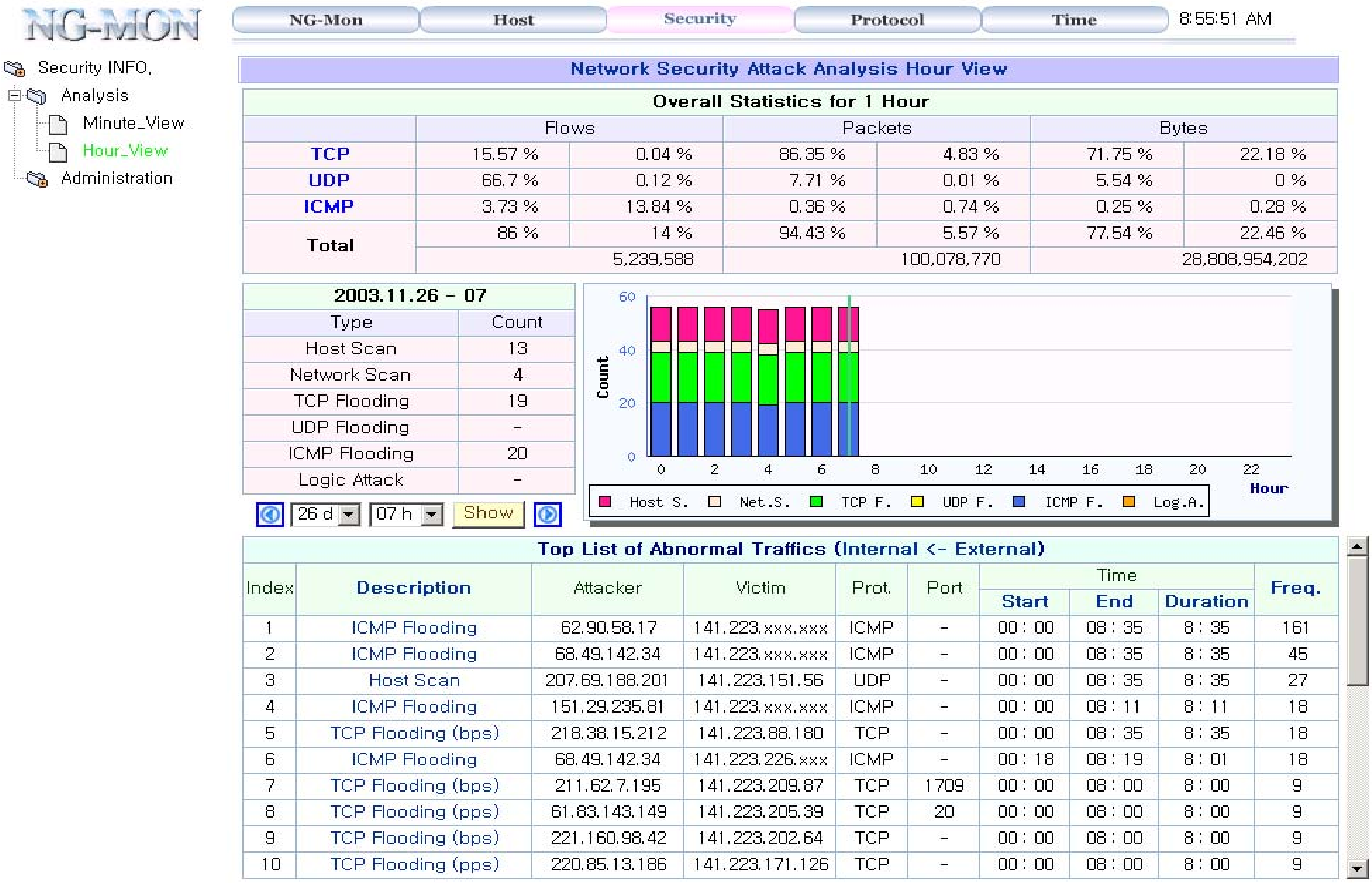
Top 10 List of Determined Applications
click_for_detail

Index	Rep. Port	Application Name	Flows		Pkts		Bytes		Detail
1	80	HTTP-WEB	11,067	6.58%	546,329	26.49%	382,643,870	27.36%	▤
2	21	FTP	335	0.2%	274,046	13.29%	245,967,147	17.59%	▤
3	9230	PDBOX	226	0.13%	146,599	7.11%	127,461,183	9.11%	▤
4	8404	V_SHARE	2,488	1.48%	134,066	6.5%	122,964,005	8.79%	▤
5	4661	eDONKEY	28,061	16.69%	193,724	9.39%	90,115,554	6.44%	▤
6	8080	HTTPS	137	0.08%	60,376	2.93%	44,105,913	3.15%	▤
7	9493	FREECHAL	92	0.05%	47,968	2.33%	42,702,255	3.05%	▤
8	22321	SORIBADA	108,773	64.68%	156,472	7.59%	28,044,089	2.01%	▤
9	1755	WMedia	135	0.08%	21,504	1.04%	26,389,585	1.89%	▤
10	6699	SAYCLUB	641	0.38%	42,424	2.06%	21,690,765	1.55%	▤

List of Undetermined Applications
click_for_detail

Index	srcPort	dstPort	P	Application Name	Flows		Pkts		Bytes		Detail
1	0	2565	T	----------	1	0%	10,987	0.53%	16,043,002	1.15%	▤
2	0	2491	T	----------	1	0%	6,310	0.31%	9,575,080	0.68%	▤
3	4459	0	T	----------	1	0%	5,341	0.26%	8,046,684	0.58%	▤
4	4887	0	T	----------	1	0%	4,917	0.24%	7,434,340	0.53%	▤
5	0	31	T	----------	1	0%	5,644	0.27%	6,529,526	0.47%	▤
6	4337	0	T	----------	1	0%	4,668	0.23%	6,514,432	0.47%	▤
7	2434	0	T	----------	1	0%	4,329	0.21%	6,465,560	0.46%	▤
8	2628	0	T	----------	1	0%	4,180	0.2%	6,194,774	0.44%	▤
9	2531	0	T	----------	1	0%	4,273	0.21%	6,094,798	0.44%	▤
10	3713	0	T	----------	1	0%	4,220	0.2%	5,462,262	0.39%	▤

NG-MON – Security Attack Analysis



Network Security Attack Analysis Hour View

Overall Statistics for 1 Hour

	Flows		Packets		Bytes	
TCP	15.57 %	0.04 %	86.35 %	4.83 %	71.75 %	22.18 %
UDP	66.7 %	0.12 %	7.71 %	0.01 %	5.54 %	0 %
ICMP	3.73 %	13.84 %	0.36 %	0.74 %	0.25 %	0.28 %
Total	86 %	14 %	94.43 %	5.57 %	77.54 %	22.46 %
	5,239,588		100,078,770		28,808,954,202	

2003.11.26 – 07

Type	Count
Host Scan	13
Network Scan	4
TCP Flooding	19
UDP Flooding	–
ICMP Flooding	20
Logic Attack	–

Top List of Abnormal Traffics (Internal <- External)

Index	Description	Attacker	Victim	Prot.	Port	Time			Freq.
						Start	End	Duration	
1	ICMP Flooding	62.90.58.17	141.223.xxx.xxx	ICMP	–	00 : 00	08 : 35	8 : 35	161
2	ICMP Flooding	68.49.142.34	141.223.xxx.xxx	ICMP	–	00 : 00	08 : 35	8 : 35	45
3	Host Scan	207.69.188.201	141.223.151.56	UDP	–	00 : 00	08 : 35	8 : 35	27
4	ICMP Flooding	151.29.235.81	141.223.xxx.xxx	ICMP	–	00 : 00	08 : 11	8 : 11	18
5	TCP Flooding (bps)	218.38.15.212	141.223.88.180	TCP	–	00 : 00	08 : 35	8 : 35	18
6	ICMP Flooding	68.49.142.34	141.223.226.xxx	ICMP	–	00 : 18	08 : 19	8 : 01	18
7	TCP Flooding (bps)	211.62.7.195	141.223.209.87	TCP	1709	00 : 00	08 : 00	8 : 00	9
8	TCP Flooding (pps)	61.83.143.149	141.223.205.39	TCP	20	00 : 00	08 : 00	8 : 00	9
9	TCP Flooding (bps)	221.160.98.42	141.223.202.64	TCP	–	00 : 00	08 : 00	8 : 00	9
10	TCP Flooding (pps)	220.85.13.186	141.223.171.126	TCP	–	00 : 00	08 : 00	8 : 00	9

5. Summary

- Internet is continuously growing in terms of: # of users & hosts, traffic loads & types
- ISPs and enterprises need to monitor their networks for various purposes (e.g., Problem Detection, Workload Characterization, Planning, SLA, Billing, Security, CRM)
- NG-MON
  - Scalable and cost-effective architecture
  - Spatial, temporal, composition analysis
  - P2P, multimedia service, game traffic analysis
  - **Network security attack analysis**

References on NG-Mon

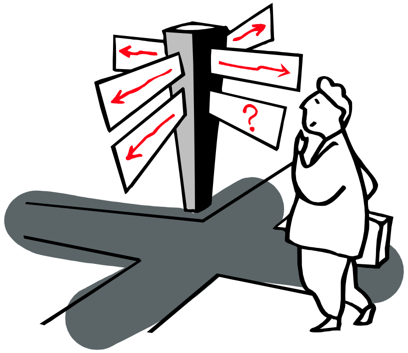
1. Myung-Sup Kim, Hun-Jeong Kang, Seong-Cheol Hong, Seung-Hwa Chung, James W. Hong, "A Flow-based Method for Abnormal Network Traffic Detection", Accepted to appear in the Proc. of the IEEE/IFIP Network Operations and Management Symposium (NOMS 2004), Seoul, Korea, April 2004.

2. Myung-Sup Kim, Hun-Jeong Kang and James W. Hong, "Towards Peer-to-Peer Traffic Analysis Using Flows", Lecture Notes in Computer Science 2867, Edited by Marcus Brunner, Alexander Keller, 14th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2003), Heidelberg, Germany, October, 2003, pp. 55-67.

3. Hun-Jeong Kang, Myung-Sup Kim and James Won-Ki Hong, "A Method on Multimedia Service Traffic Monitoring and Analysis", Lecture Notes in Computer Science 2867, Edited by Marcus Brunner, Alexander Keller, 14th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management (DSOM 2003), Heidelberg, Germany, October, 2003, pp. 93-105.

4. Hun-Jeong Kang, Seung-Hwa Chung, Seong-Cheol Hong, Myung-Sup Kim and James W. Hong, "Towards Flow-based Abnormal Network Traffic Detection", Proc. of 2003 Asia-Pacific Network Operations and Management Symposium (APNOMS 2003), Fukuoka, Japan, October 1-3, 2003, pp. 369-380.

Questions?



jwkhong@postech.ac.kr

http://dpnm.postech.ac.kr/ (my research lab)

http://ngmon.postech.ac.kr (NG-Mon)