

# Measurement of Data Traffic in Cellular Networks

2008.8.16

Kisu Kim, Hyeongu Son, Taek-keun Kwon, DK Lee\*, S. Moon\* and Youngseok Lee,

Chungnam National University

\*KAIST

Daejon, Korea

# Contents

- Overview
- Basic measurement result
- Result of content classification
- Summary

# 3G Cellular Data Networks in Korea

- 3G data networks
  - GSM/UMTS (3GPP) family
    - UMTS/W-CDMA (3G): HSDPA (High-Speed Downlink Packet Access)
  - CDMA (3GPP2) family
    - CDMA-2000 (3G): EVDO (Evolution Data Only/Optimization)
  - Provides always-on connectivity services
    - In a mobile environment: even in the express train at the speed of 300Km/H

	<b>CDMA 1x EVDO</b>	<b>WCDMA/HSDPA</b>
Uplink	153.6Kbps	2Mbps
Downlink	2.4Mbps	14.4Mbps

# 3G Subscribers in Korea



Source: IDATE

[http://www.giikorea.co.kr/press/IU57012\\_kr.html](http://www.giikorea.co.kr/press/IU57012_kr.html)



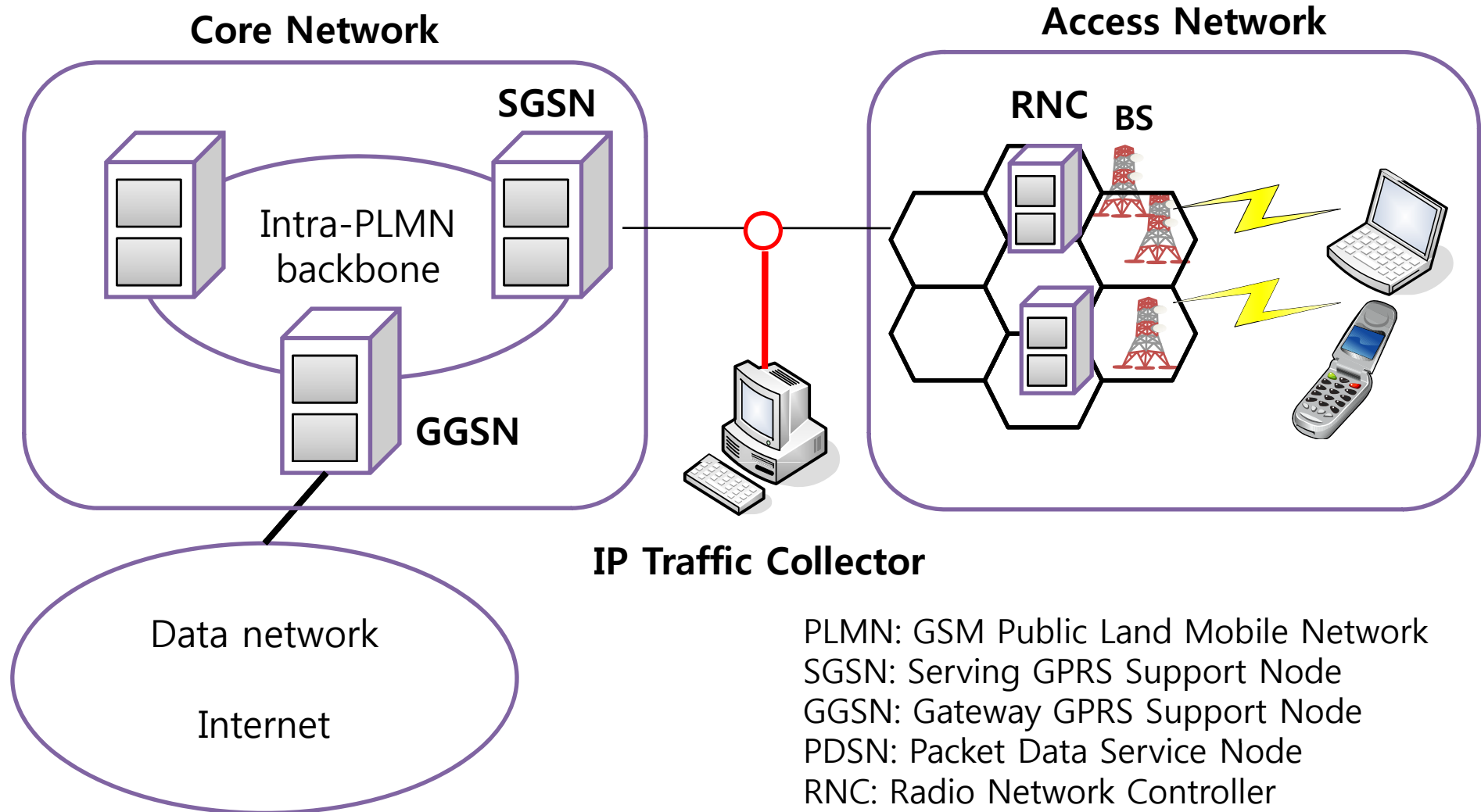
# Our Work

- Measurement and classification of EVDO/HSDPA data traffic
- To answer the questions like
  - What is the trend of protocol and port breakdown ?
    - Similar in the typical wired Internet ?
  - Which kind of contents are people using ?

# Collected Traffic

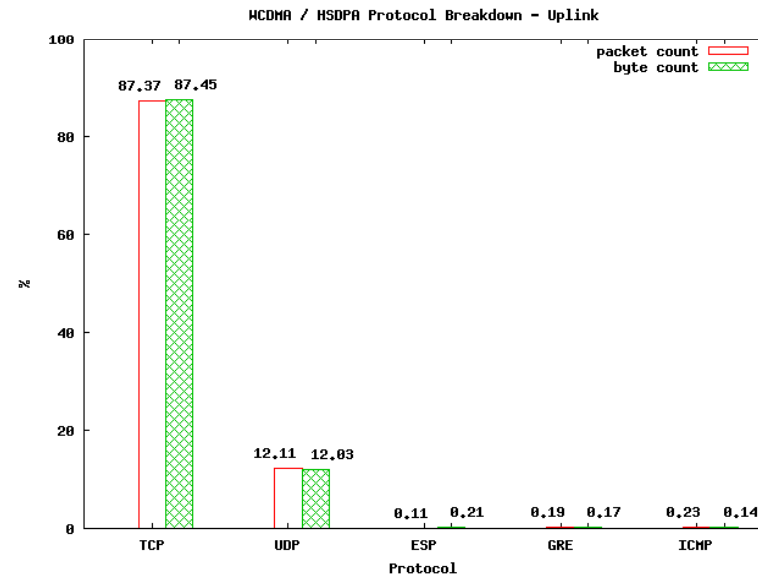
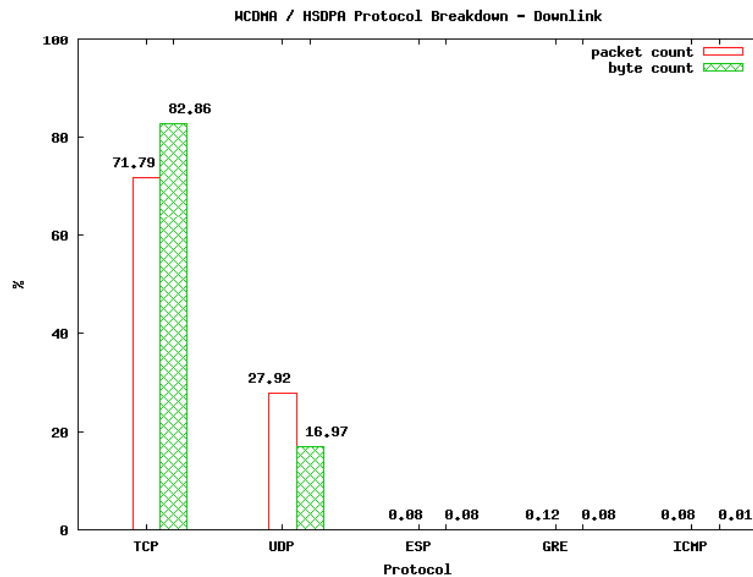
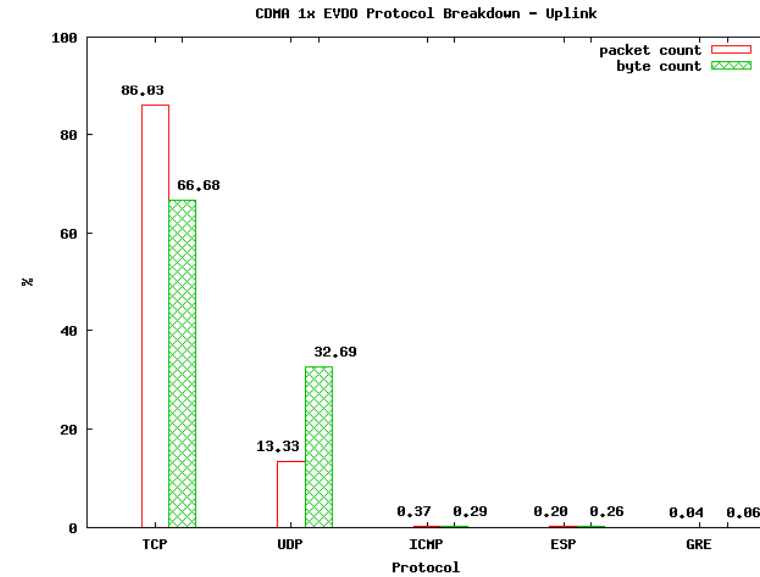
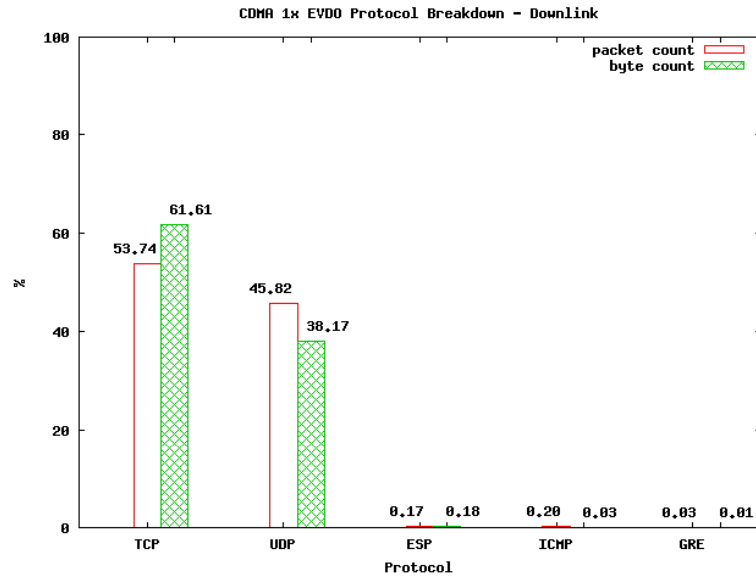
- One of major cellular carriers in Korea
- CDMA 1x EVDO
  - Eight 30-min traces on different days in 2008
- WCDMA HSDPA
  - Eight 30-min traces on different days in 2008

# Measurement Environment



PLMN: GSM Public Land Mobile Network  
SGSN: Serving GPRS Support Node  
GGSN: Gateway GPRS Support Node  
PDSN: Packet Data Service Node  
RNC: Radio Network Controller  
BS: Base Station

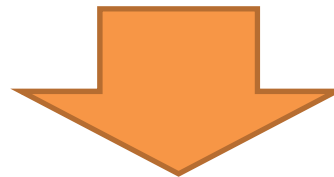
# Protocol Breakdown





# Protocol Breakdown (cont'd)

	CDMA 1x EVDO		WCDMA / HSDPA	
	Download	Upload	Download	Upload
TCP (%)	Byte : 61.6 Pkts : 53.7	Byte : 66.7 Pkts : 86.0	Byte : 82.9 Pkts : 71.8	Byte : 87.5 Pkts : 87.4
UDP (%)	Byte : 38.1 Pkts : 45.8	Byte : 32.7 Pkts : 13.3	Byte : 17.0 Pkts : 27.9	Byte : 12.0 Pkts : 12.1



RTP/UDP (byte %)	95.19	3.89	87.78	0.68
RTP/Total (byte %)	35.41	1.05	15.32	0.07

# Port Breakdown

Rank	CDMA 1x EVDO				WCDMA / HSDPA			
	Download (Src. Port)		Upload (Dst. Port)		Download (Src. Port)		Upload (Dst, port)	
	TCP	UDP	TCP	UDP	TCP	UDP	TCP	UDP
1	9093	15000	9093	1813	80	6970	80	1814
2	80	15006	8082	5060	9093	15004	9093	4500
3	7231	15004	80	15321	7231	4500	443	10001
4	7237	15002	12100	53	554	15006	8082	10000
5	7236	15008	443	13771	443	5004	20	137

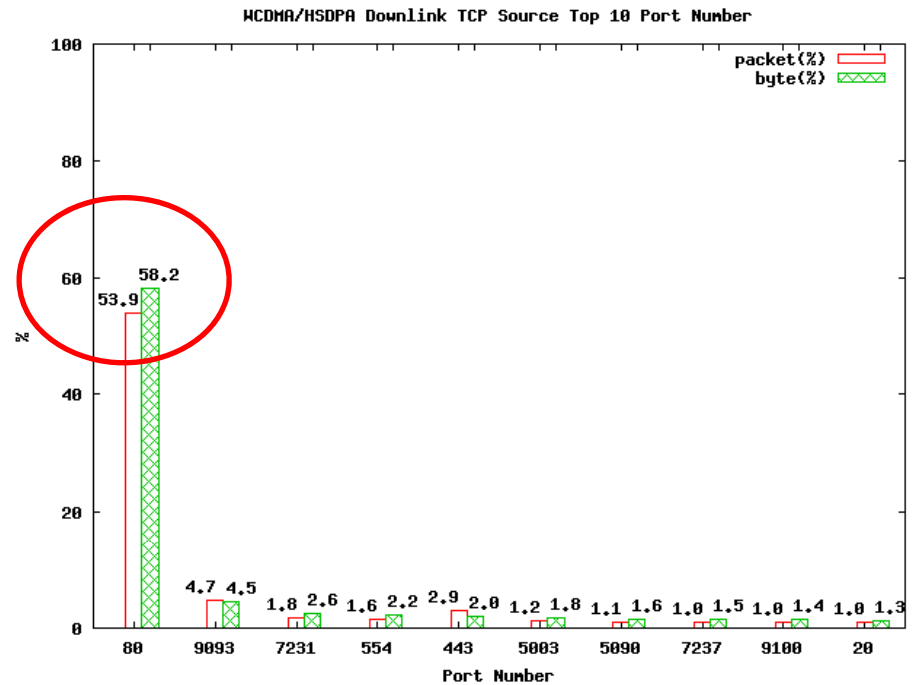
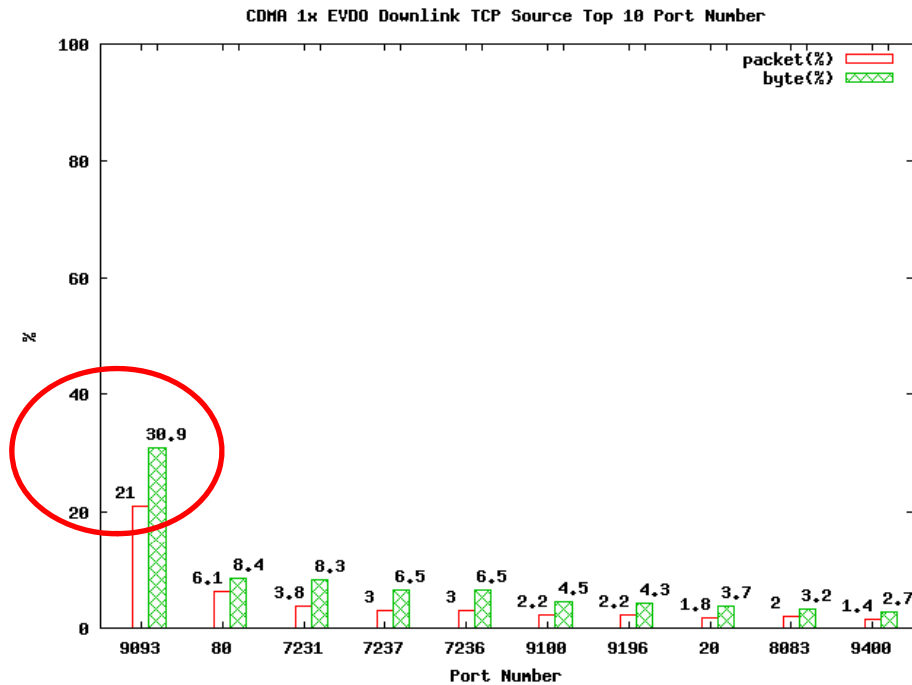
9093 : WAP protocol over HTTP

80 : HTTP

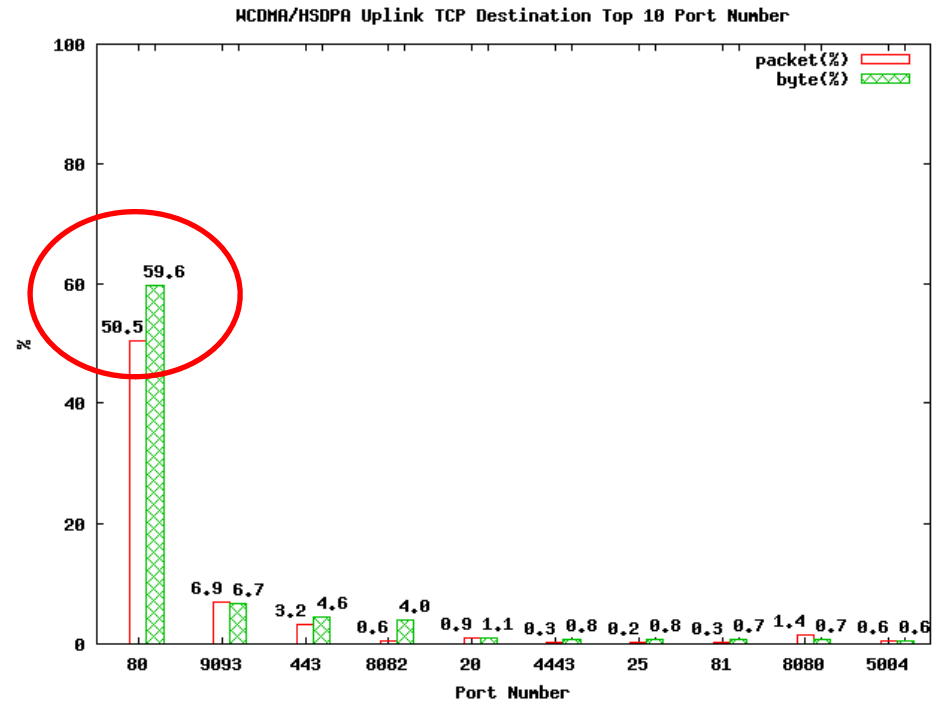
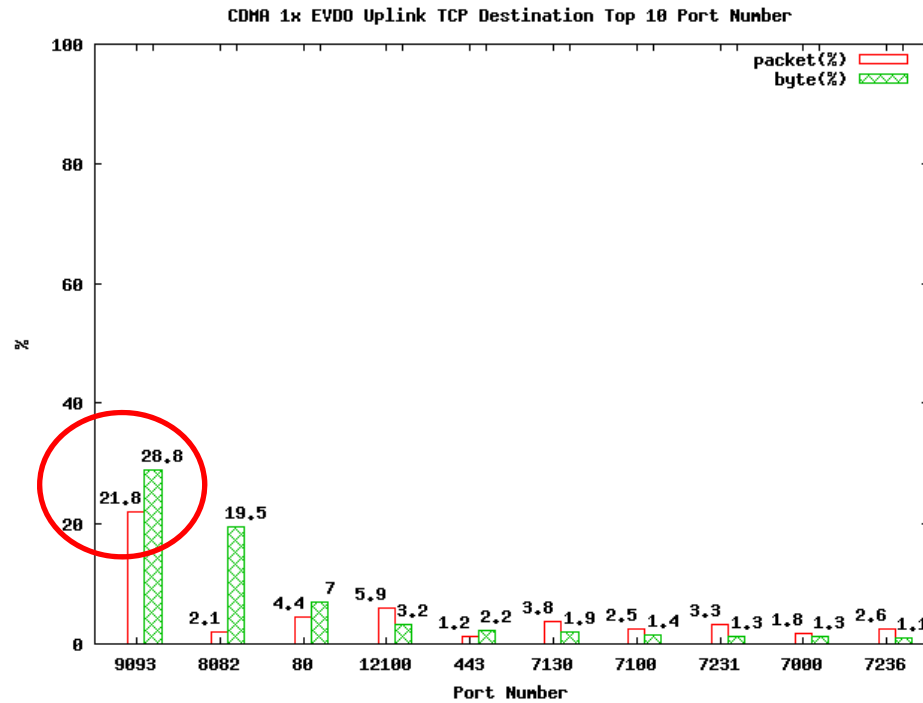
7231 : authentication for downloading music

Around 15000 : for listening to music

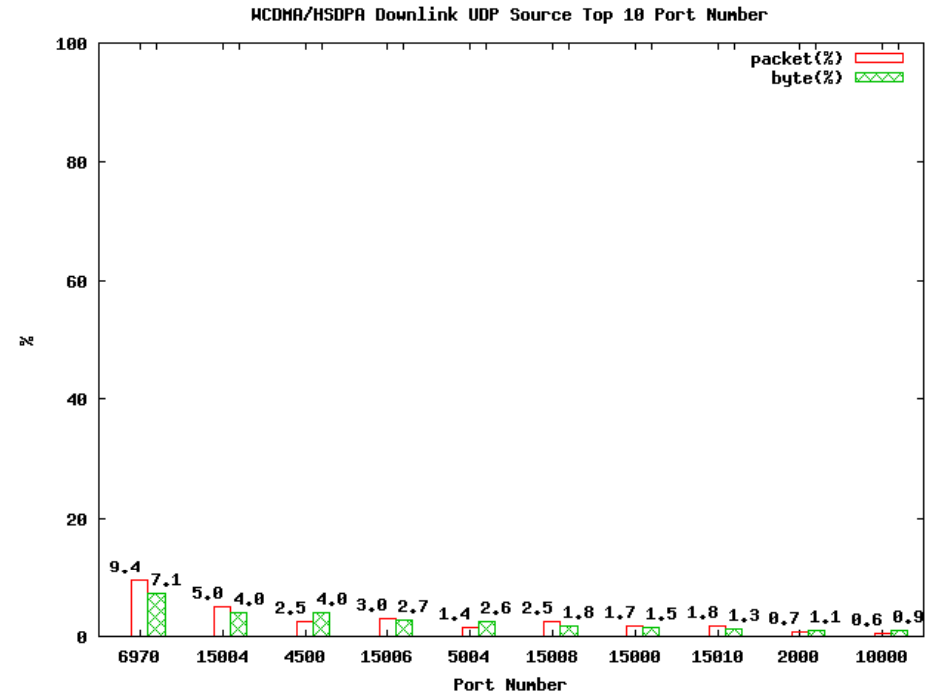
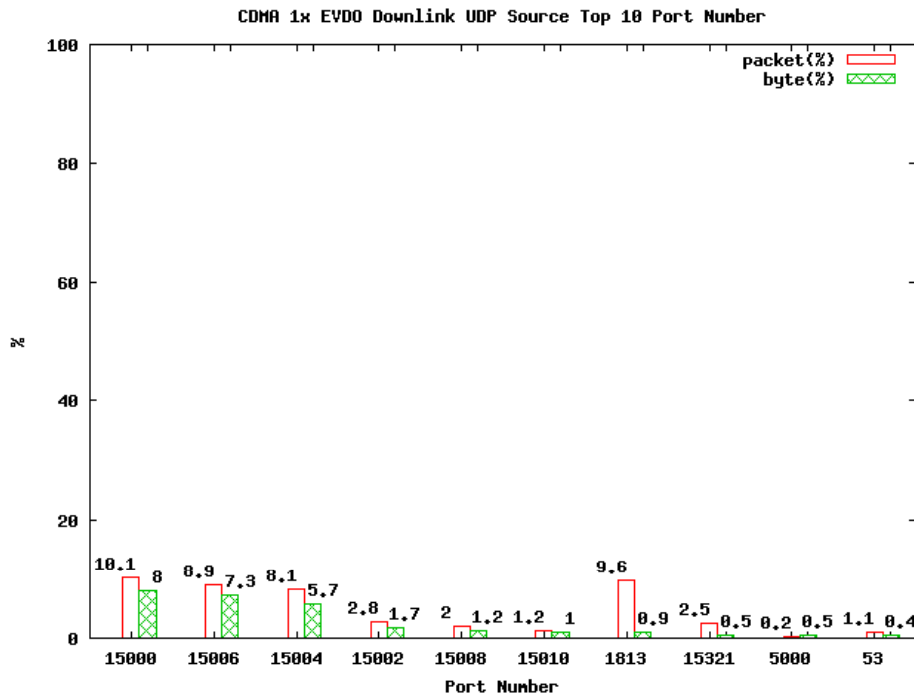
# TCP Src Port - Downlink



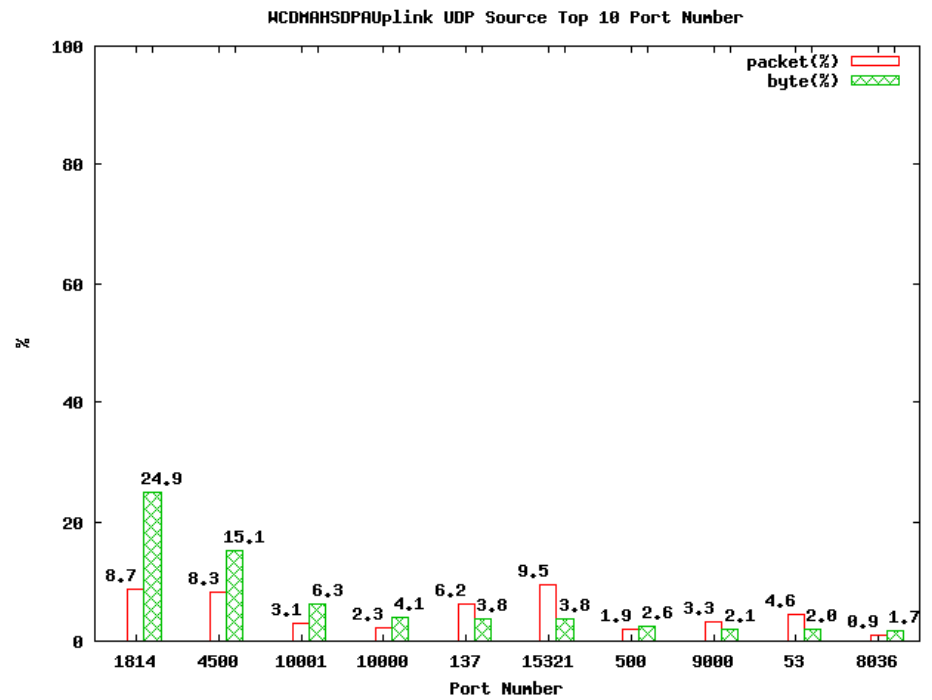
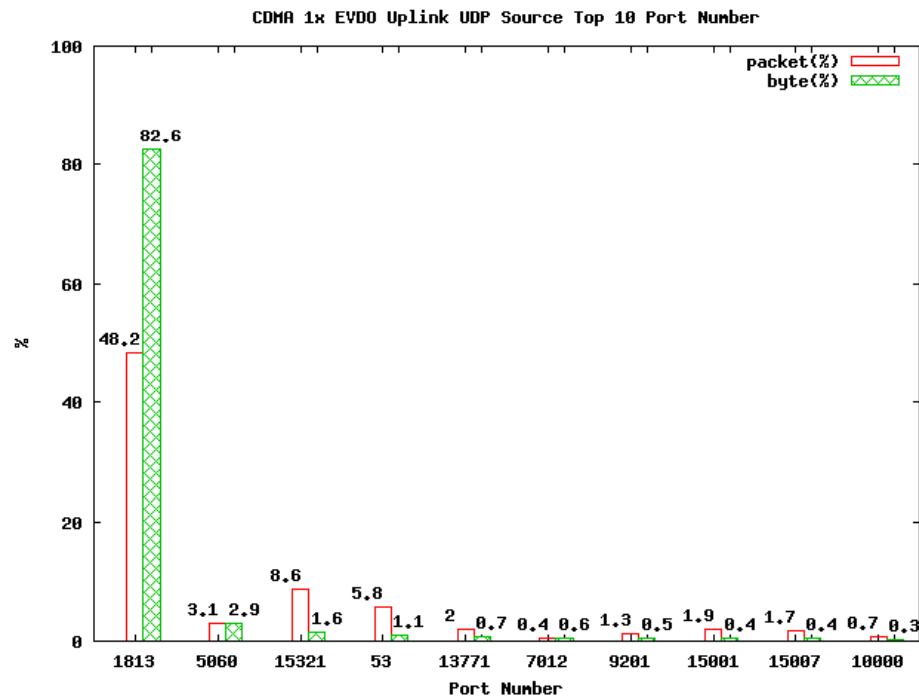
# TCP Dst Port – Uplink



# UDP Src Port - Downlink

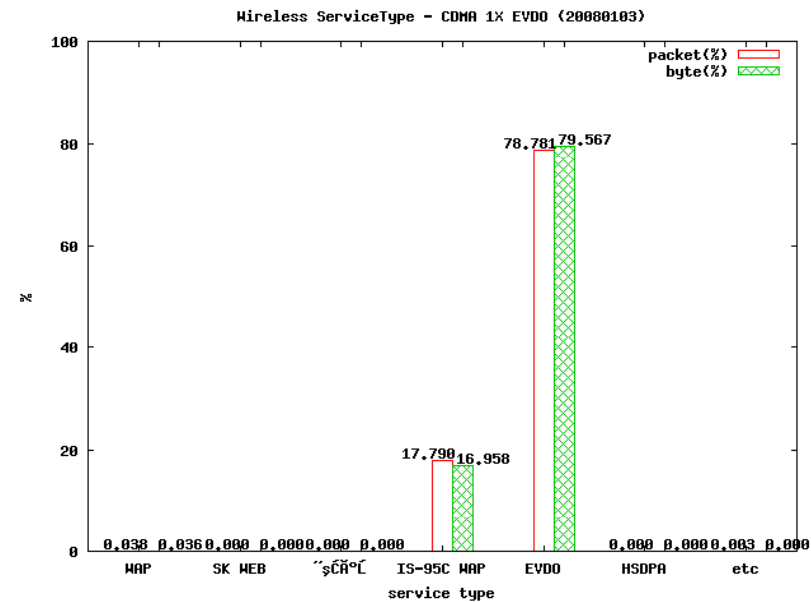
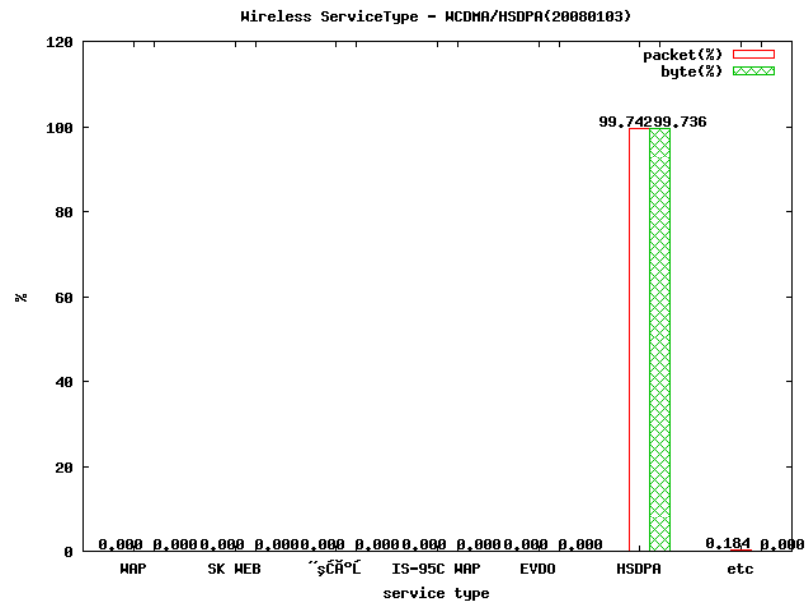


# UDP Dst Port - Uplink



# Interesting Port: 9093

- Port 9093
  - WAP over HTTP
  - Classification of cellular “service types” in HTTP request packets



# Port 80

- Mobile terminals access to web via port 80 ?
  - Mostly by PCs not by cellular phones
- Classification of terminal types with "agent" in HTTP packets

	Byte count(%)	Packet count(%)
Laptops/PCs	66	69
Cellular phone	0.006	0.009



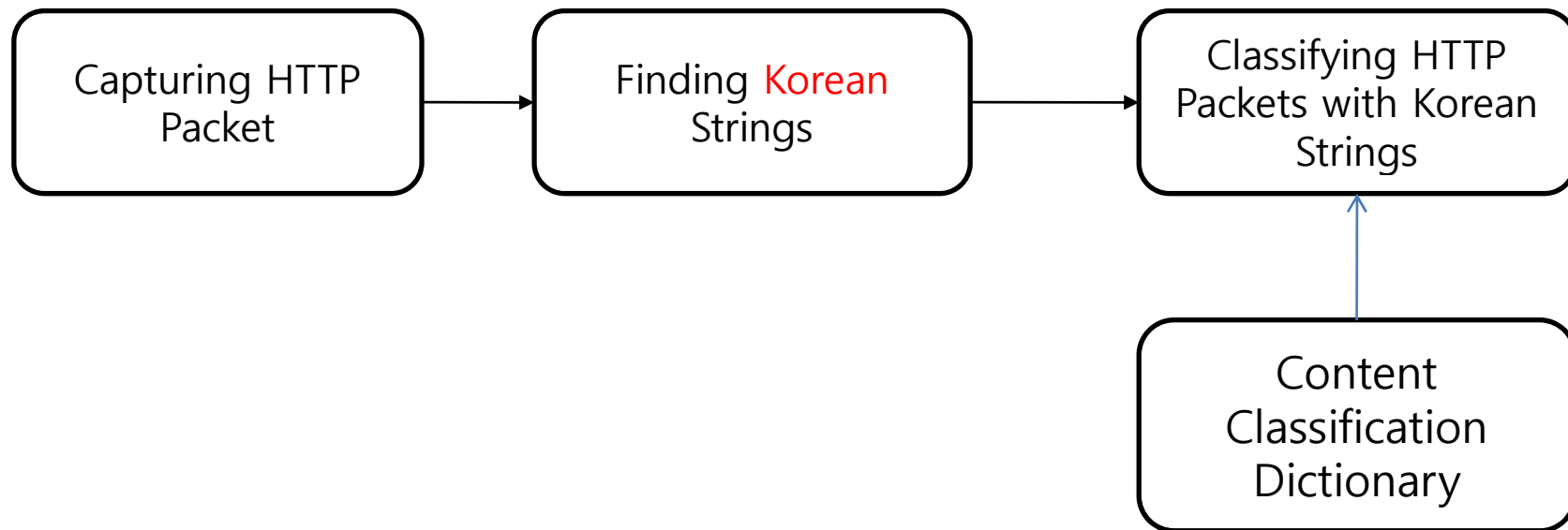
# Content Classification of HTTP Traffic

- For Internet traffic classification
  - Traditional methods that use
    - Port number
    - Payload signatures
    - Behaviors/Machine learning
  - Classification of traffic: web, p2p, email, ftp
- Content classification of HTTP traffic
  - E.g.) Web traffic is 40%. Which contents are people using with web traffic ?
    - Searching popular keywords, reading news, shopping

# How to Classify Contents of HTTP Traffic ?

- Analyze servers' log data
  - Easy
  - but limited to
    - Only content providers
- Inspect packets
  - Difficult
    - *May have privacy issues*
  - ISPs as well as content providers
    - Even campus network operators can classify the campus networks' Internet usage pattern
  - What do we find by inspecting HTTP packets ?
    - Specific "strings" that provide hints for classification
    - We need language-specific information: A lot of contents are localized

# Overview of Content Classification for HTTP Packets



Korean Strings	Content Category
"news", name of news company, ...	News
"*.mp3", name of songs and singers, ...	Music



# Typical Korean Encoding

- **KS C 5601**
  - KS X 1001, 0xB0A1(가) ~ 0xC8FE(힉)
  - 2 Bytes, Korean character set – 2350 syllables
- **EUC-KR**
  - KS X 1003(≡ ASCII) + KS X 1001
  - 2 Bytes
- **Unicode :**
  - U+AC00(가) ~ U+D7A3(힉), Korean character set -- 11172
- **UTF-8**
  - U+0800 ~ U+FFFF : 3 Byte encoding
- **UTF-16**
  - U+0800 ~ U+FFFF : 2 Byte encoding

# Example of Finding Korean String

```
Hypertext Transfer Protocol
HTTP/1.1 200 OK\r\n
Request Version: HTTP/1.1
Response Code: 200
Date: Thu, 03 Jan 2008 03:09:09 GMT\r\n
Server: Microsoft-IIS/6.0\r\n
X-Powered-By: ASP.NET\r\n
Expires: Tue, 01 Feb 2000 00:00:00 GMT\r\n
Content-Length: 9132\r\n
Content-Type: text/vnd.wap.wml\r\n
charset:=KS_C_5601_1987/r/n
Cache-control: no-cache\r\n
\r\n
extensible Markup Language
<?xml
version="1.0"
encoding="ks_c_5601-1987"
?>

0710 30 30 30 30 30 30 05 01 0c 01 72 5a 25 30 30 30 0000,co for.#000
0720 30 30 30 3b 74 65 78 74 2d 64 65 63 6f 72 61 74 000;text -decorat
0730 69 6f 6e 3a 6e 6f 6e 65 72 3e 0d 0a 5b bc ad bf ion:none "> [...
0740 ef bd c5 b9 ae 5d b9 e0 c0 ba 20 bb f6 20 c1 f8 ....]
0750 b9 d9 c1 f6 bf cd 20 bc d5 b9 d9 b4 da b8 b8 20 .....
0760 c7 d1 20 c7 da b5 e5 b9 e9 2c 20 be e3 c0 ba 20 .....
0770 ba a7 c6 ae bf cd 20 b1 ea c5 d0 20 b8 d3 b8 ae .....
0780 20 c0 e5 bd c4 bf eb c7 b0 a1 a6 2e 20 b8 da c0 .....
0790 ef c0 cc b6 f3 b8 e9 20 bf c3 c7 d8 20 b4 ab bf .....
07a0 a9 b0 dc ba c1 be df 20 c7 d2 20 3c 61 20 68 72 ..... <a hr
07b0 65 66 3d 22 68 74 74 70 3a 2f 2f 74 6d 65 74 61 ef="http //tmeta
07c0 2e 6e 61 74 65 2e 63 6f 6d 3a 37 30 30 35 2f 69 .nate.co m:7005/i
07d0 6d 63 2e 64 6f 3f 72 73 64 3d 52 53 44 30 30 30 mc.do?rs d=RSD000
07e0 30 30 30 31 26 61 6d 70 3b 72 6d 64 3d 26 61 6d 0001&amp ;rmd=&am
07f0 70 3b 6b 65 79 77 6f 72 64 3d 78 74 43 38 78 77 p;keywor d=xtC8xw
```

[서울신문] 밝은 색 진  
바지와 손바닥만 한  
핸드백, 얇은 벨트와  
깃털 머리 장식용품....  
멋쟁이라면 올해 눈여  
겨보아야 할

0xBCAD 서  
0xBFEB 울  
0xBDC5 신  
0xB9AE 문  
0xB9E0 밝  
0xC0BA 은  
0xBBF6 색  
:

# Examples of Classifying HTTP Packets by Packet Inspection

- Contents delivered in HTTP packets
  - Listening to music
  - Reading news
  - Browsing stock information

# Listening to Music with Cellular Phone in HTTP

```

Frame 14 (1518 bytes on wire, 1518 bytes captured)
Ethernet II, Src: Cisco_22:82:c0 (00:13:80:22:82:c0), Dst: StarentN_02:0e:31 (00:05:47:02:0e:31)
Internet Protocol, src: 211.234.229.25 (211.234.229.25), Dst: 10.136.90.101 (10.136.90.101)
Transmission Control Protocol, src Port: 9093 (9093), Dst Port: 32866 (32866), Seq: 0, Ack: 442, Len: 1460
  
```

```

01e0 0d 0a 0d 0a 0d 0a 0d 0a 0d 0a 09 3c 77 6d 6c 3e
01f0 0d 0a 09 09 3c 63 61 72 64 20 69 64 3d 22 6d 65
0200 6e 75 22 20 6e 65 77 63 6f 6e 74 65 78 74 3d 22
0210 74 72 75 65 22 20 74 69 74 6c 65 3d 22 b0 cb bb
0220 f6 b0 e1 b0 fa 20 3a 20 5b 33 2f 36 5d 22 20 3a
0230 0d 0a 09 09 20 20 20 20 3c 70 3e 20 b0 cb bb f6
0240 be ee 20 3a 20 6b 63 6d 20 3c 62 72 2f 3e 0d 0a
0250 09 09 09 3c 61 20 68 72 65 66 3d 22
0260 73 65 61 72 63 68 2e 6a 73 70 3f 73 6e 32 3d 6b
0270 63 6d 26 61 6d 70 3b 73 74 32 3d 26 61 6d 70 3b
0280 63 74 49 64 32 3d 30 33 36 30 31 26 61 6d 70 3b
0290 53 4d 3d 30 31 30 30 30 30 30 26 61 6d 70 3b 53
02a0 55 3d 30 30 30 32 34 35 37 36 38 33 5f 30 30 30
02b0 30 30 30 30 30 30 30 5f 43 30 30 30 30 30 30 30
02c0 31 33 26 61 6d 70 3b 50 43 3d 22 3e c0 e7 b0 cb
02d0 bb f6 3c 2f 61 3e 3c 2f 70 3e 0d 0a 09 20 20 20
02e0 20 09 09 3c 62 72 2f 3e 0d 0a 09 09 09 09 0d
02f0 0a 09 20 20 20 20 20 20 20 20 09 09 3c 64 6f 20
0300 74 79 70 65 3d 22 76 6e 64 2e 73 6b 6d 6e 31 22
0310 0d 0a 09 20 20 20 20 20 20 20 20 09 09 09 6e 61
0320 62 65 6c 3d 22 c0 cf b3 e2 c0 fc 20 bf c0 b4 c3
0330 2d 4b 43 4d 22 3e 0d 0a 09 20 20 20 20 20 20
0340 20 09 09 3c 67 6f 20 68 72 65 66 3d 22 68 74 74
0350 70 3a 2f 2f 6a 75 6e 65 70 61 67 65 2e 6e 61 74
0360 65 2e 63 6f 6d 3a 31 30 31 2f 70 70 2e 6a 73 70
0370 3f 70 61 67 65 5f 49 44 3d 30 30 33 37 33 35 35
0380 37 38 36 26 61 6d 70 3b 53 4d 3d 30 31 30 30 30
0390 30 30 26 61 6d 70 3b 53 55 3d 30 30 30 32 34 35
03a0 37 36 38 33 5f 30 30 30 30 30 30 30 30 30 5f
03b0 43 30 30 30 30 30 30 31 33 26 61 6d 70 3b 50
  
```

```

...<car d id="me
nu" newc ontext="
true" ti tle="...
.... : [3/6]" >
... <p> ...
... : kcm <br/>...
  
```

card id="menu"  
 newcontext="true"  
 title="검색결과 : [3/6]"    검색어 :  
 kcm

```

.. <do
type="vnd.skmn1"
... ..la
bel="... ..
-KCM">...
..<go href="htt
p://june page.nat
e.com:10 1/pp.jsp
?page_ID =0037355
786& SM=01000
00& S U=000245
7683_000 0000000_
c0000000 13& P
  
```

do  
 type="vnd.skmn1" lab  
 el="일년전 오늘-KCM

# Stock Information

Ethernet II, Src: Cisco\_22:85:86 (00:13:86:22:85:86), Dst: Starlink\_02:0e:31 (00:03:47:02:0e:31)  
 Internet Protocol, Src: 211.234.229.25 (211.234.229.25), Dst: 10.143.128.65 (10.143.128.65)  
 Transmission Control Protocol, Src Port: 9093 (9093), Dst Port: 34543 (34543), Seq: 5841, Ack: 535, Len: 1460

0010	05	dc	6d	1b	40	00	3d	06	89	2c	03	ea	e5	19	0a	81	..m.0.;. . . . .
0020	80	41	23	85	86	ef	d8	7b	e9	95	28	7c	12	17	50	18	.A#....{ ..( ..P.
0030	ff	ff	3b	64	00	00	33	35	33	38	33	22	20	61	63	63	..;d..35 383" acc
0040	65	73	73	6b	65	79	3d	22	34	22	20	73	74	79	6c	65	esskey=" 4" style
0050	3d	22	6d	61	72	71	75	65	65	2d	73	74	79	6c	65	3a	="marque e-style:
0060	73	6b	74	73	63	72	6f	6c	6c	3b	20	63	6f	6c	6f	72	sktscrol l; color
0070	2a	20	23	30	30	30	30	30	30	22	3e	53	4b	c1	f5	b1	: #00000 0">SK...
0080	c7	2a	3c	2f	61	3e	0a	20	20	20	20	20	20	20	20	20	.*</a>.
0090	20	20	20	20	20	20	20	20	20	20	20	3c	2f	74	64	3e	</td>
00a0	0a	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	</tr>
00b0	20	3c	2f	74	72	3e	20	20	0a	20	20	20	20	20	20	20	<tr>
00c0	20	20	20	20	20	20	20	20	20	3c	74	72	3e	0a	20	20	<td c1 ass="td_
00d0	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	orange ce 1" widt
00e0	20	20	3c	74	64	20	63	6c	61	73	73	3d	22	74	64	5f	h="50%"> 12:38(
00f0	6f	72	61	6e	67	65	5f	63	65	31	22	20	77	69	64	74	...)</td >.
0100	68	3d	22	35	30	25	22	3e	20	31	32	3a	33	38	28	c7	<td
0110	f6	c0	e7	29	3c	2f	74	64	3e	0a	20	20	20	20	20	20	class= "td_oran
0120	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	ge_ce2 f c_blue
0130	64	20	63	6c	61	73	73	3d	22	74	64	5f	6f	72	61	6e	align_c" width="
0140	67	65	5f	63	65	32	20	66	63	5f	62	6c	75	65	20	20	50%">3,7 15.</td
0150	61	6c	69	67	6e	5f	63	22	20	77	69	64	74	68	3d	22	<
0160	35	30	25	22	3e	33	2c	37	31	35	bf	f8	3c	2f	74	64	<tr>
0170	3e	0a	20	20	20	20	20	20	20	20	20	20	20	20	20	20	<td cla ss="td_o
0180	20	20	3c	2f	74	72	3e	0a	20	20	20	20	20	20	20	20	range ce 2 fc blu
0190	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	e align_c">..90
01a0	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	..</td>.
01b0	20	3c	74	64	20	63	6c	61	73	73	3d	22	74	64	5f	6f	<td
01c0	72	61	6e	67	65	5f	63	65	32	20	66	63	5f	62	6c	75	class="t d_orange
01d0	65	20	20	61	6c	69	67	6e	5f	63	22	3e	a1	e5	39	30	ce2 fc blu
01e0	bf	f8	3c	2f	74	64	3e	0a	20	20	20	20	20	20	20	20	<td
01f0	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	class="t d_orange
0200	63	6c	61	73	73	3d	22	74	64	5f	6f	72	61	6e	67	65	ce2 fc blu
0210	5f	62	65	72	70	66	62	5f	62	6c	75	65	20	20	61	6e	

SK증권\*

12:38(현재)

3,715원

▼90원



# Reading News

☒ Frame 45 (1518 bytes on wire (1194 bytes captured) on interface 0: Ethernet II)	
☒ Ethernet II, Src: Cisco_22:82:c0 (00:13:80:22:82:c0), Dst: StarentN_02:0e:31 (00:05:47:02:0e:31)	
☒ Internet Protocol, Src: 211.234.229.25 (211.234.229.25), Dst: 10.137.207.132 (10.137.207.132)	
☒ Transmission Control Protocol, Src Port: 9093 (9093), Dst Port: 32928 (32928), Seq: 10740, Ack: 869, Len: 1460	

01b0	61 74 69 6f 6e 3a 6e 6f 6e 65 22 3e 0d 0a 5b bc	ation:ne" >... [
01c0	ad bf ef bd c5 b9 ae 5d b9 e0 c0 ba 20 bb f6 20	.....] .....
01d0	c1 f8 b9 d9 c1 f6 bf cd 20 bc d5 b9 d9 b4 da b8	.....
01e0	b8 20 c7 d1 20 c7 da b5 e5 b9 e9 2c 20 be e3 c0	.....
01f0	ba 20 ba a7 c6 ae bf cd 20 b1 ea c5 d0 20 b8 d3	.....
0200	b8 ae 20 c0 e5 bd c4 bf eb c7 b0 a1 a6 2e 20 b8	.....
0210	da c0 ef c0 cc b6 f3 b8 eb 20 bf c3 c7 d8 20 b4	.....
0220	ab bf a9 b0 dc ba c1 be df 20 c7 d2 20 3c 61 20	..... <a
0230	68 72 65 66 2d 22 68 74 74 70 2a 2f 2f 74 6d 65	href="ht tp://tm
0240	74 61 2e 6e 61 74 65 2e 63 6f 6d 3a 37 30 30 35	ta.nate. com:7005
0250	2f 69 6d 63 2e 64 6f 3f 72 73 64 3d 52 53 44 30	/imc.do? rsd=RSD0
0260	30 30 30 30 30 31 26 61 6d 70 3b 72 6d 64 3d 26	000001&a mp;rmd=&
0270	61 6d 70 3b 6b 65 79 77 6f 72 64 3d 78 74 43 38	amp;keyw ord=xtC8
0280	78 77 25 33 44 25 33 44 26 61 6d 70 3b 72 75 6c	xw%3D%3D &amp;ru1
0290	3d 61 48 52 30 63 44 6f 76 4c 32 35 33 63 79 35	=aHR0cDo vL253cy5
02a0	77 63 33 6c 75 5a 58 51 75 59 32 38 75 61 33 49	wc31uzXQ uY28ua3I
02b0	36 4f 54 41 35 4d 43 39 30 62 32 49 79 4c 33 52	60TA5MC9 0b2IyL3R
02c0	76 59 69 39 6f 62 33 51 76 61 47 39 30 4d 6a 42	vY19ob3Q vaG90mJB
02d0	33 4d 44 49 75 59 58 4e 77 50 31 4e 57 51 30 51	3MDIuYXN wP1NwQ0Q
02e0	39 4d 6a 59 6d 51 30 64 53 50 54 45 6d 51 55 6c	9MjYmQ0d SPTEmQU1
02f0	45 57 44 30 78 4d 7a 6b 77 4f 44 41 78 4a 6b 6c	EWd0xMzk wODAxJk1
0300	4e 52 30 35 50 50 54 45 6d 53 55 31 48 53 55 52	NR05PPTE msU1HSUR
0310	59 50 54 41 6d 55 45 35 50 50 54 45 6d 56 30 59	YPTAMUE5 PPTEmV0Y
0320	39 4d 54 41 3d 26 61 6d 70 3b 54 49 4d 45 4f 55	9MTA=&am p;TIMEOU
0330	54 3d 26 61 6d 70 3b 53 55 3d 30 30 30 32 38 33	T=&amp; S U=000283
0340	36 36 38 30 26 61 6d 70 3b 53 4d 3d 30 30 30 30	6680&amp ;SM=0000
0350	30 30 30 22 20 73 74 79 6c 65 3d 22 66 6f 6e 74	000" sty le="font
0360	2d 63 6f 6c 6f 72 3a 23 66 66 35 36 30 33 3b 74	-color:# ff5603;t
0370	65 78 74 2d 64 65 63 6f 72 61 74 69 6f 6e 3a 6e	ext-deco ration:n
0380	6f 6e 65 22 3e c6 d0 bc c7 3c 2f 61 3e 20 be c6	one">... </a> ..

[서울신문] 밝은 색 진바지와 손바닥만한 핸드백, 얇은 벨트와 깃털 머리 장식용품.... 멋쟁이라면 올해 눈여겨봐주세요

# Experimental Results

- HTTP packet traces
  - CDMA 1x EVDO
  - WCDMA HSDPA
  
- Port 9093 in HTTP

# Content Classification of Port 9093 Traffic

CONTENTS CATEGORIES		EVDO		HSDPA	
		Access Count %	Packet %	Access Count %	Packet %
Entertainment	Music/Video Streaming/Downloading	12.54	14.01	15.03	15.52
	Adult	0.42	0.18	0.47	0.18
	Game	10.24	5.47	10.16	6.06
Commerce	Shopping	2.54	2.96	3.30	2.89
	Stocks	4.29	6.20	2.73	2.29
Information	News & Web	25.66	27.54	24.26	29.38
	Community	7.23	7.32	7.46	7.24
	Customer service	20.13	11.15	31.35	16.52
Public information	Police enforcement information	1.18	4.57	0.15	0.03
	Transportation information (bus map)	0.59	0.22	0.53	0.20
	Public advertisement	0.88	0.29	0.16	0.03
N.A.	N.A.	1.41	20.08	4.40	19.65

# Summary

- Data traffic in cellular networks
  - Dominant TCP
  - Popular UDP/RTP streaming
- Content classification
  - Language-specific HTTP traffic classification
    - Useful for finding user behaviors of using Web services
- Future work
  - Supporting several Korean codes
  - Automatically building a classification dictionary
  - Experiments with various HTTP packet traces
  - Extension with other languages