A Study on Smart-phone Traffic Analysis

Sang-Woo Lee, Jun-Sang Park, Hyun-Shin Lee, Myung-Sup Kim
Dept. of Computer and Information Science
Korea Univ.
Yeongi, Korea
{sangwoo_lee, junsang_park, hyunshin-lee, tmskim}@korea.ac.kr

Abstract—The increase in Smart-phone users and expansion of market value creates traffic complexity and causes network saturation. Network technology lags development compared to the growth of traditional internet applications and Smart-phone based applications. Thus, the need for interest in network traffic monitoring increase. Traffic monitoring becomes an important element in the management of stable, efficient network. This paper analyzes various results of Smart-phone traffic in the campus network. It compares the characteristics and differences among traditional internet traffic and Smart-phone traffic.

Keywords—Smart-phone; traffic characteristic; traffic analysis; campus network; application traffic classification;

I. INTRODUCTION

The appearance and rapid development of Smart-phones changes our environment, including changing the paradigm of industrial structures. Network communication is changing rapidly due to the development of Smart-phones. In an open platform environment, the development of new types of applications, due to the diversification of user needs and differentiated quality of service requirements are increasing. The network provider wants to minimize network management costs and to provide high quality services using new revenue streams.

Smart-phone traffic has increased by more than 300 times over the past two years due to the explosive increase in Smart-phone devices and subscribers that join an unlimited data payment system and changes of usage patterns. The author’s of [1] predict Smart-phone data traffic will increase 30~40 times in the next three years. [1]

Vendors are expanding the bandwidth of their network equipment to ensure network stability and reliability to cope with the increased traffic network. However, the performance of network equipment is limited.

In this situation, the network manager needs to understand the high resource load consumed by services. This requires management policies to guarantee network performance through Traffic-Engineering, Network-Planning, QoS-Planning, SLA using service monitoring of the subscribers’ patterns.

The importance of network traffic analysis is increasing due to the growth of Apple Appstore [5] and Android open market places [6] and the appearance of various Smart-phone services and the development of application programs.

This paper shows the overall characteristics of Smart-phone traffic using a trace collected in a campus network. Flow is the analysis viewpoint and shows network layer characteristics. This reveals differences between traditional internet traffic and Smart-phone traffic characteristics.

The remainder of the paper is organized as follows. Section 2 describes problems of existing research in Smart-phone traffic analysis, and our Smart-phone traffic classification methodology. Section 3 describes our experimental environment and traffic traces. Section 4 describes the distribution and composition of Smart-phone traffic over the entire traffic network. Section 5 describes Smart-phone application traffic analysis. Section 6 describes characteristics of Smart-phone traffic compared to traditional internet traffic. Section 7 draws conclusions and outlines future work.

II. RELATED WORK

Hossein Falaki’s research describes smart device traffic characteristics, based on application usage frequency, using the time per subscriber. [2] They collected raw information for the traffic analysis using a self-developed program from end terminals (Smart-phone) because they found it hard to classify Smart-phone device traffic in the variety of wired and wireless traffic in the coexisting network environment using techniques, such as tabs or mirroring.

However, this method of collecting traffic for the entire target network is difficult to apply to other network users. Only limited information can be collected from small populations, and such traffic is likely to distort the results.

Gregor Maier’s research described NAT usage and estimated the number of devices under NAT in residential broadband networks. [3] They used IP TTL and HTTP user-agent string information to count the number of devices based on OS and browser information.
However, they provided only the lower bound of device numbers and cannot provide information if devices do not generate HTTP traffic.

In this study, traffic was collected on the top of the target network link. Smart-phone traffic was classified using [4]'s methods for accurate, reasonable analysis. Our previous study [4] used the header information of packets to identify the traffic’s terminal operating system. The previous study focused attention on the packet header field values to classify the host’s operating system. Using Nop Count, Timestamp, SACK, Initial TTL, window size scale, window scale factor, window size, maximum segment size, EOL, etc. We probed unique packet header fields per operating system. Table 1 describes the classification of the version of the Smart-phone operating systems that inspect the packet header. In this study, we extract Smart-phone traffic from the trace. Then, we show the traffic characteristic and distribution, based on that methodology.

Table 1 describes the identifiable Smart-phone operating system.

<table>
<thead>
<tr>
<th>Smart-phone Device</th>
<th>OS Version</th>
<th>iPhone</th>
<th>Android</th>
<th>Windows Mobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS Version</td>
<td>iOS 3.x</td>
<td>iOS 4.x</td>
<td>Cupcake</td>
<td>WinMo 6.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Éclair</td>
<td>WinMo 6.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Froyo</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Gingerbread</td>
<td></td>
</tr>
</tbody>
</table>

Gregor Maier’s research [5] described smart device traffic analysis based on the application layer protocol. The research showed that 57–79% of traffic was formed with a web based application and HTTP protocol. However, this protocol criteria analysis result cannot show detailed information to the network manager. Thus, in this study we show the result based on application payload signature classification results.

### III. EXPERIMENTAL ENVIRONMENT

In this section, we describe the system environment to extract Smart-phone traffic in the enterprise network. We describe the traffic trace’s information used in this paper.

Figure 1 illustrates the experimental environments configured in the campus network and externally connected to the Internet to collect packets from the top-level router. Packets are transformed into the flow and delivered to the analyzer. The analyzer identifies the generated traffic’s host operating system based on packet header information. It marks the operating system information in the specific fields. The Smart-phone traffic classifiers extract Smart-phone traffic using those fields.

Table 2 describes the entire traffic information collected for the experiment. The trace was formed with public PC, laptop, Web servers, including Smart-phone traffic. About four thousand hosts generated traffic with various applications (e.g. P2P, Web, and Instant messenger).

<table>
<thead>
<tr>
<th>Time</th>
<th># of Flows</th>
<th># of Packets</th>
<th>Bytes(TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>$8 \times 10^8$</td>
<td>$389 \times 10^8$</td>
<td>340</td>
</tr>
<tr>
<td>Portion</td>
<td>2.36 %</td>
<td>3.87 %</td>
<td>4.12 %</td>
</tr>
</tbody>
</table>

Figure 2 describes the proportion of Smart-phone traffic over time. The high proportions are for Byte, Packet, and Flow, sequentially. The maximum proportion for Byte is 40.1%.
Packet 34.2%, Flow 8.2%. The highest percentages of time shown were at 10 AM and 2 PM. The network traffic usage changes express the lifestyle patterns of network members.

Figure 2 shows a high byte proportion for March 12 (Saturday), March 13 (Sunday) but the proportion averages 5%. The reduction in the total amount of traffic status is due to the lab always using wired traffic, and the school officials do not go to work in the weekend.

Figure 2. Proportion of Smart-phone traffic over time among total traffic

Figure 3 describes the byte proportion of Smart-phone traffic. We can illustrate the highly non-ideal bytes proportion measurement in Figure 2 through Figure 3. This is due to the total bytes in the weekend decreasing by more than ten times compared to the weekday peak, especially, at March 12 near 10 AM, and March 13 near 0 AM.

Figure 3. Amounts of bytes of Smart-phone, total traffic

B. Smart-phone host distribution

Figure 4 show the Smart-phone host distribution over time. The daily task time starts at 10 AM; the variation in the host distribution is greatest at that time. The proportion of the operating systems is for the iPhone, Android, and Windows Mobile, respectively. The maximum count of the operating system was iPhone 176, Android 14, Windows Mobile 17.

Table 4 shows the maximum user count, average user count per minute. The average number of minutes for a host was for the iPhone, confirming it had the largest number of users.

<table>
<thead>
<tr>
<th>MAX</th>
<th>iPhone</th>
<th>Android</th>
<th>Windows Mobile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Per Minutes</td>
<td>27.88</td>
<td>1.23</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Figure 4. Distribution of Smart-phone devices based on OS

C. Distribution of Smart-phone traffic Traditional Protocol

Figure 5 shows the traditional protocol traffic proportion among Smart-phone traffic.

31.9% of the HTTP traffic flow accounted for the largest proportion; for the DNS it was 24.1%, HTTPS had 6.6%, and SSH had 1.2%. However, in bytes, HTTP had 50%, HTTPS had 1.5%, and DNS had 0.2%. This shows that the HTTP protocol generated the most traffic. That means the many Smart-phone application use HTTP protocol, and web based.
Figure 5. Distribution of Smart-phone traffic using Traditional Protocol

D. Smart-phone HTTP traffic analysis

Application service providers use various ports to bypass the firewall to provide users with seamless service. For example, traditional HTTP protocol uses port 80, but they use another port for HTTP. Thus, this section indicates the distribution of HTTP server ports. Figure 6 shows the server port distribution for Smart-phone HTTP traffic. Port 80 had most, 97.86%, of the distribution and 3228, 7080, 8080, 9090, 15151 ports used small proportions. Port 45800 showed large variation in packet, bytes; it was identified as file transfer via HTTP protocol.

Figure 6. Server port distribution of Smart-phone HTTP traffic

V. SMART-PHONE APPLICATION TRAFFIC ANALYSIS

In this section, we present the Smart-phone application traffic classification result. The analysis target application showed iPhone apple Appstore [6], Android Market [7], in the top ranking 50 apps. We make their payload signature to classify application traffic. And we classified the traffic using the extracted payload signature. Table 5 shows classification completeness based on the payload signature classifier machine in the pre classified Smart-phone traffic [8].

That classification result shows the completeness of Smart-phones traffic. Flows completeness is 15.37%, Packet 10.52%, Bytes are 7.70%. We classify Smart-phone application traffic apple Appstore, Android Market in the top 50 ranking apps. However, the results show less than 20% completeness. This denotes users of Smart-phone use a variety of applications. Surveys on the applications for the Smart-phone traffic analysis should be performed continuously. We are considering an automated payload signature generation system focused on Smart-phone application using LCS.

<p>| TABLE V. COMPLETENESS OF SELECTED SMART-PHONE APPLICATION BASED ON SIGNATURE |
|-------------------------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Completeness</th>
<th>Flow %</th>
<th>Packet %</th>
<th>Bytes %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness</td>
<td>15.37%</td>
<td>10.52%</td>
<td>7.70%</td>
</tr>
</tbody>
</table>

Table 6 describes the top ten Smart-phone applications based on the proportion of bytes. The iPhone app Appstore generates the most traffic in the trace. That is, many iPhone users use application updates and download new applications via this application. The web browsing application safari, korea message exchange application kakaotalk (similar in us, whatsapp), streaming application youtube, and social network application cyworld, and navigation application navermap, follow in decreasing order.

| TABLE VI. BYTE PROPORTION OF TOP 10 APPLICATIONS |
|-----------------------------------------------|-----------------|-----------------|-----------------|
| Rank                                         | Application Name | Flows | Packets | Bytes | Bytes Proportion (%) |
|-----------------------------------------------|-----------------|-----------------|-----------------|
| 1                                             | Appstore        | 129K         | 37M         | 42GB   | 12.51                |
| 2                                             | safari          | 828K         | 29M         | 22GB   | 6.19                 |
| 3                                             | kakaotalk       | 329K         | 7M          | 1.53GB | 0.41                 |
| 4                                             | youtube         | 10K          | 1M          | 1.41GB | 0.39                 |
| 5                                             | cyworld         | 29K          | 927K        | 632MB  | 0.18                 |
| 6                                             | navermap        | 10K          | 404K        | 293MB  | 0.08                 |
| 7                                             | Facebook        | 27K          | 507K        | 234MB  | 0.07                 |
| 8                                             | daummap         | 6K           | 208K        | 129MB  | 0.04                 |
| 9                                             | nateon          | 14K          | 421K        | 125MB  | 0.03                 |
| 10                                            | twitter         | 13K          | 348K        | 124MB  | 0.03                 |
VI. CHARACTERISTICS OF SMART-PHONE TRAFFIC COMPARED TO TRADITIONAL INTERNET TRAFFIC

A Smart-phone is smaller than existing endpoint hardware and software, due to the mobility of the endpoint. Thus, Smart-phone traffic has distinct characteristics compared to traditional Internet traffic. This section describes the characteristics of Smart-phone traffic compared to traditional Internet traffic.

A. Flow duration

Figure 7 shows the flow duration distribution. Most Smart-phones flow durations are shorter than those for traditional internet traffic. Less than 50% of the Smart-phone’s flow is short, less than 1 second duration. This identified almost HTTP traffic and application keep-alive packets for the sustained connection, and polling. Figure 7 shows 45% of the flow has less than 2 seconds duration for non-Smart-phone traffic, identifying most P2P application’s search function traffic.

B. Packets per Flow distribution

Figure 9 shows packets per flow graph. The graph shows Smart-phone packets per flow was much greater than for non-Smart-phone’s.

The number of communications with less than ten packets occupies 70% of the flow in the non-Smart-phone traffic. This reveals the P2P application, identified by the searching packets. Smart-phone traffic that has less than ten packets per flow occupies 60% of the flow, because most Smart-phone applications use HTTP protocol for small web content downloads.

Before reversion of Smart-phone duration compared to non-Smart-phone traffic, less than 3 seconds, is identified that login, authentication, script download for Smart-phone application. After reversion, that is identified real web browsing, application data, data download using dispersion connection mechanism.
C. Bytes per packet distribution

Figure 11 shows the bytes per packet distribution. In general, non-Smart-phone applications generate much larger packets. 80% of non-Smart-phone traffic is less than 320 bytes; that for Smart-phones is less than 380 bytes.

Figure 11. Distribution of bytes per packet

Figure 12 describes the HTTP packets bytes per packet distribution. 10% of Smart-phone’s packets are approximately 128 bytes. These are identified as sending or receiving short sentence messaging application packets. Less than 320-byte traffic is identified as real web browsing data optimized web pages for Smart-phone devices. Traffic exceeding 320 bytes is identified as data downloaded via HTTP protocol, transferring streaming data.

![Bytes Per Packet](image)

Figure 12. Distribution of HTTP bytes per packet

VII. SUMMARY

The average proportion of Smart-phone traffic in the trace is 2~4%; the maximum measurement value is 40% in terms of bytes, explaining why traffic decreased on Saturday and Sunday. Most Smart-phone users use iPhone. The user count, in average per minute, for iPhone is 27.9, for Android is 1.2, and Windows Mobile is 1.9. HTTP protocol of Smart-phone’s traffic occupied the most bytes. DNS protocol accounts for most of the flow. This shows that most of Smart-phone applications are based on the HTTP protocol. Most Smart-phone HTTP traffic uses port 80, with another 2% using another port.

We classify the Smart-phone application traffic based on the payload signature and classifier. However, the low completeness means each user uses a variety of applications.

Smart-phone’s flow duration is short compared to that of traditional Internet traffic. Smart-phone’s applications have many more packets per flow than does traditional internet traffic. However, the traffic is almost the same for HTTP. 80% of the traffic has less than 320 bytes. However, the proportion for HTTP protocol is almost the same.

VIII. CONCLUSIONS AND FUTURE WORK

This paper analyzed traffic generated by Smart-phone in a campus network. It indicated Smart-phone traffic compared to non-Smart-phone traffic characteristics. Smart-phone’s traffic have characteristic that compared with non-Smart-phone traffic. We can consider the characteristic design for a Smart-phone traffic application classification system.

We will consider other Smart-phone traffic characteristics in future work. We plan to specify other Smart-phone application traffic characteristics. E.g. Traffic pattern of famous applications. We will develop a method of automated
signature generation system for efficient Smart-phone application traffic classification.

REFERENCES


