# Traffic Characteristics of Modern Video Streaming Service

Jinsoo Yoo, Junyeop Kim, Youjip Won, Hoyoung Kim, Seungki Park, and Taenyeon Ha

***Abstract*—The network traffic caused by mobile devices has been growing. Moreover, N-screen service which enables to watch video contents in various network devices accelerates the traffic growth. To handle the high network traffic effectively, the service provider should grasp the characteristics of the traffic. In this paper, we make a minute observation about the traffic of N-screen service using network traces of iShoot program which provide such a service. From the in-depth study, we find that there is a specific traffic pattern repeated every day. Also we observe contents popularity distribution and the traffic imbalance between the servers, and so on. These results can be used to design N-screen service architecture or maintain quality of video streaming service in the other mobile services.**

***Index Terms*—N-screen, traffic analysis, video streaming service, etc.**

## I. Introduction

Network technologies like Cloud-Computing and N-screen enable video contents to be subsequently played on various mobile devices. There are many video streaming service programs support N-Screen service, and iShoot [1] published by IUMedia Company is one of the N-screen video streaming service programs. Recently, however, encouraged by a great development in the wireless communication technology, the traffic of internet services caused by mobile devices has been rapidly increased. To maintain streaming service quality and control the mobile network traffic effectively, we have to know the characteristics of the traffic. Although there are many traffic analysis of traditional video streaming services, N-screen Video sharing services like iShoot are different from past video streaming services in several aspects. First, in the past video streaming services, a particular contents provider uploads contents to the streaming server and manages them, and clients just download the contents. On the other hand, in the N-screen video streaming services, unspecified individuals actively upload their contents to a cloud server and use it. Because of this, it is hard for the service provider to know the number of new contents which will be uploaded in a day. Second, N-screen video streaming service changes the pattern of traffic generated by network devices. By using N-screen video streaming service, Network traffic is generated from not a PC but also smartphone, mp4 player and many other network devices, and people can play video contents anywhere, anytime in their mobile device. From these facts, we can infer that the properties of N-screen service traffic are different from the other traditional video streaming services.

Manuscript received March 17, 2012; revised May 5, 2012.
J. Yoo, J. Kim, and Y. Won are with the Dept. of Electronics and Computer Engineering, Hanyang University, Seoul, Korea. (e-mail: jedisty@hanyang.ac.kr; tel.: +82-2-2220-4579; fax: +82-2-2220-4579.).
H. Kim, S. Park, and T. Ha are with the IUMedia Company, Seoul, Korea.

To analyse N-screen service traffic, we gathered Apache logs generated by the video streaming servers which work for iShoot N-screen service. Each Apache log contains information about when the clients sent the request, how many bytes were transferred to client, which contents were transferred, and so on. There is one Apache server log per a client streaming request. We used Apache logs collected from iShoot N-screen service. Using these traces, we examined closely the various aspects of N-screen service traffic. From the results, we find some notable traffic characteristics which may help service providers expect server traffic and organize the streaming servers.

The rest of this paper is structured as follow. Section 2 describes how iShoot provide N-screen video service to clients and how we used Apache log to analyse traffic. In section 3, we examined the N-Screen service traffic from various angles and also considered how the traffic can be handled effectively. Section 4 describes other studies about traffic analysis. Finally, we summarize and conclude this paper in Section 5.

## II. IShoot N-Screen Video Service and Traffic Analysis

### A. How IShoot Works

Users can easily make use of N-screen video streaming service through iShoot program. If someone wants to play video contents on the mobile device, he or she should upload the contents to the iShoot server, and then, he can play the video contents on his mobile devices through iShoot program. More details on this process and how iShoot servers are co-worked are shown in Fig.1

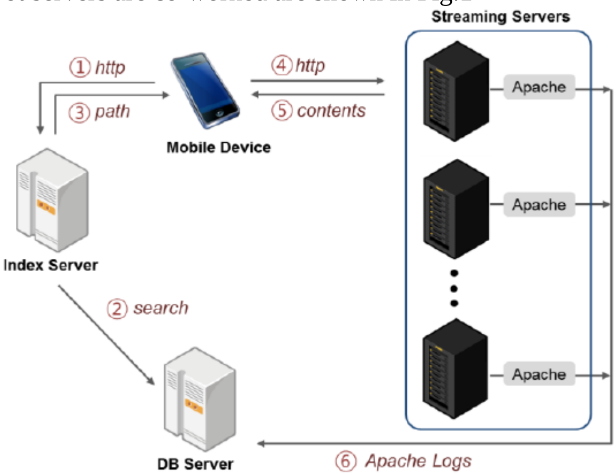



Fig. 1. iShoot service architecture

Video contents uploaded by clients are saved in streaming servers. After the contents are saved, the process that a client gets the specific video contents is as follows: First, a video streaming request from a client is sent to Index server by http protocol. Second, Index server looks up DB Server to get the contents path of the contents. The path informs which streaming server has the video contents, and DB Server maintains the path information. Third, Index server hands the contents path to the client's device. Fourth, the client sends a request to the streaming server which has the video contents. Then, the streaming server sends the video contents to the client. Finally, at the end of each request completion, apache program operating on each streaming servers makes a log, Apache log, containing information about when the clients sent the request, how many bytes were transferred to client, which contents were transferred and so on. This will be discussed in details in next section.

### B. *Apache Logs*

TABLE.I: ISHOOT APACHE LOG INFORMATION

| Item | Description |
|---|---|
| Record Time | The completion time of a transmission for a client request |
| Client IP | IP address to identify clients |
| Contents | File path in where the contents is saved |
| Transferred Byte (byte) | The number of bytes transferred to a client for a request |
| Service Time (us) | The transmission time for a request |
| User Agent | The identifying information that the client browser reports about itself |
| Result | Response code |

When a client watches video contents using iShoot program, the Streaming server transfers the video contents to the clients. With each transmission is finished, the information about the transmission is saved in the form of Apache log. The information contained in Apache log is shown in Table I. Noting that the Record Time in Apache log denotes the when this request was completed not when this request arrived to iShoot. The request arrival time can be calculated by subtracting the service time from record time. We manipulate the data in Apache logs to analyse characteristics of iShoot N-screen service.

## III. ANALYSIS RESULT

To study traffic characterization in N-Screen video service, we collected a considerable amount of Apache logs generated from 16 streaming servers by apache web server program. The days we collected the logs differ from each streaming server. The log collection periods for each server are shown in Table II.

### A. *Traffic Pattern*

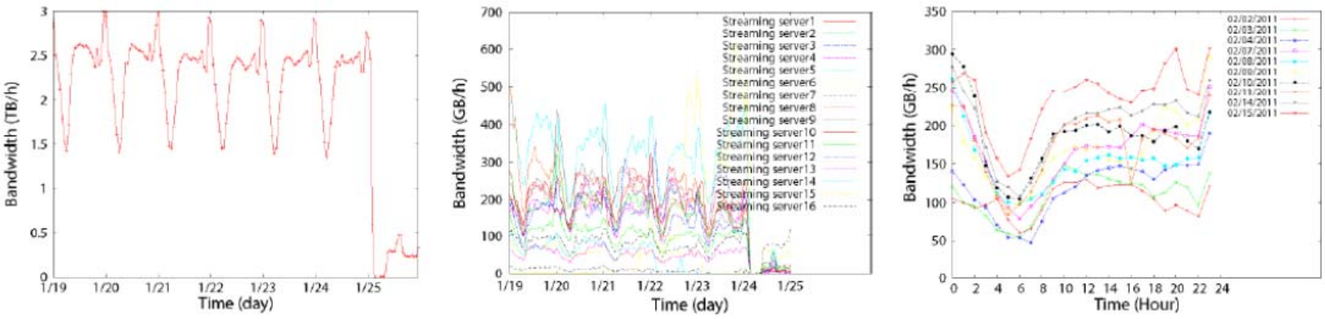
We observe daily traffic pattern for each streaming server and daily total traffic pattern for all streaming servers. To obtain the data, first we make per-hour bandwidth data for each streaming server. The bandwidth data can be gained by summing transferred byte in Apache log which generated at the same hour. We use Apache log dates from 01/19/2011 to 01/25/2011. Daily total traffic pattern is shown in Fig. -(a) and daily traffic pattern for each server is shown in Fig. -(b). Additionally, we observe the bandwidth flow in a day of streaming server 10 using Apache log from 02/09/2011 to 02/19/11. It is shown in Fig. -(c).

TABLE II: LOG COLLECTION PERIODS

| Server | Log Collection Period |
|---|---|
| Streaming server 1 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 2 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 3 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 4 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 5 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 6 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 7 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 8 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 9 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 10 | 01.01.2011 ~ 09.05.2011 |
| Streaming server 11 | 01.01.2011 ~ 09.05.2011 |
| Streaming server 12 | 01.01.2011 ~ 09.05.2011 |
| Streaming server 13 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 14 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 15 | 01.01.2011 ~ 02.25.2011 |
| Streaming server 16 | 01.01.2011 ~ 02.25.2011 |

In Fig. 2-(a) and Fig. 2-(b), X-axis is date and Y-axis is bandwidth (TB/h). Fig. -(a) show that there is a specific traffic pattern repeated every day. The bandwidth is peak at about 10 p.m. to 1 a.m. every day and has the lowest value at about 4 a.m. to 7 a.m. Fig. -(c) displays the traffic pattern in a day more clearly and shows the fact that this pattern is maintained every day. The pattern is certainly bound up with the client's life pattern. This means the people actively use n-screen video streaming service at the end of the day.

The detail distribution of the bandwidth across the streaming servers is shown in Fig. -(b). From the graph, we can make the following observation. Similarly with Fig. 2-(a), each streaming server has a daily bandwidth patterns. But, noting that, the patterns look different from each other. The time when each streaming server has a peak bandwidth does not coincide with each other. The each streaming server has own traffic pattern and a peak traffic time in a day. It means that the number of requests is quite different from each streaming servers at certain times. We call the contents accessed frequently hot contents, and this phenomenon is caused from the skewing of the hot contents distribution across the streaming servers. In N-screen video sharing service, favourite video contents are uploaded in and out of season and requests for the video contents from the other users can be increased steeply. Such behaviour can leads to imbalance distribution of hot contents among streaming servers. To prevent this phenomenon and spread high traffic over the servers, the hot contents which are just uploaded to streaming server are properly distributed over some streaming servers.



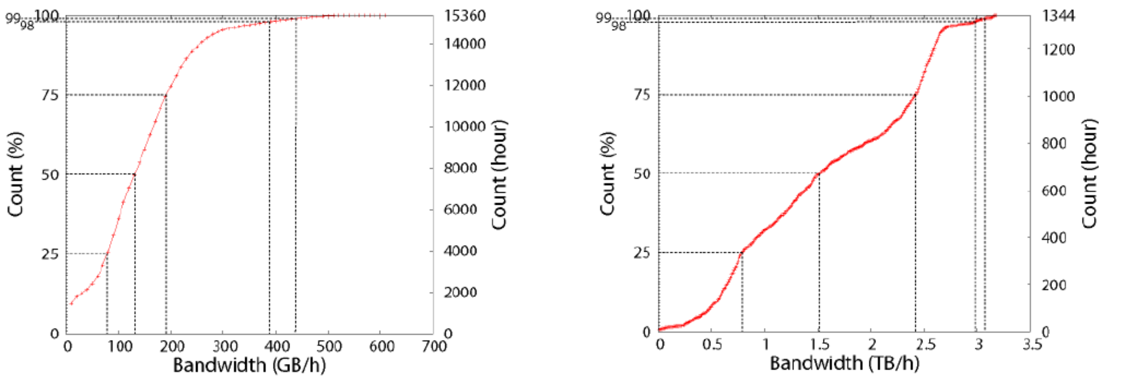

(a) Total bandwidth of all streaming servers (b) Bandwidth of each streaming server (c) Bandwidth of streaming sever 10

Fig. 2. Daily traffic distribution.

### B. *Traffic Distribution*

How well streaming servers worked together can be judged by how well traffic is distributed through the servers. So we observe traffic distribution of all streaming servers as follow. First, the bandwidth, traffic per hour, is calculated using Apache logs for each server. This denotes the total traffic transferred from each streaming server during each one-hour and we used Apache logs from 01/01/2011 to 02/25/2011 for all streaming servers. We make histogram of that with 10GB bin size using all streaming server bandwidth data. Finally, we draw CDF graph with the histogram. This graph is shown in Fig. 3-(a). For in-depth observation, we make another graph in different way. First, we calculate the bandwidth, traffic per hour, using Apache logs for each server and made total bandwidth by aggregating the bandwidths occurred same time and date. This data denotes the total traffic transferred from all streaming servers during each one-hour. Then we make histogram of that with 10GB bin size and draw CDF graph with the histogram. This graph is shown in Fig. 3-(b).

As explained, Fig. 3 -(a) and Fig. 3-(b) are made using same Apache logs. But, interestingly, the shape of the two graphs is quite different. Fig. 3-(a) which is the CDF for sum of each histogram is skewed. In the figure, the portion of low bandwidth is bigger than the high bandwidth portion. This implies that the streaming servers used many times to transfer data in low bandwidth. In contrast, CDF for aggregated bandwidth histogram increases more linearly than Fig. 3-(a). From this observation, we can infer that, actually, there were many hours that the streaming servers transferred contents at high aggregated bandwidth. And it is opposite to Fig. 3-(a). This difference is caused by the traffic imbalance between the streaming servers. The reason why Fig. 3-(a) is skewed toward low bandwidth is that a few streaming servers treated high traffic load while many other streaming servers maintained low bandwidth at the same time, so the portion of low bandwidth is bigger than high bandwidth portion.




(a) CDF of a summation of histogram for each streaming server

(b) CDF of a histogram of aggregated traffic for a total streaming server

Fig. 3. CDF of a histogram for streaming server traffic

### C. *Contents Popularity Distribution*

To confirm the popularity distribution over contents, we calculate the bytes transferred to clients in a day for each contents. The observation is conducted with 10 days Apache logs of streaming server 10. We rank the traffic and convert it into a percentage CDF. Fig. shows the result. In Fig. , X-axis denotes a percentage of contents and Y-axis denotes accumulated traffic, and each line represents the one day popularity distribution. From the graph, we can confirm several characteristics for the contents popularity distribution. First, the popularity distribution is similar among each day. It implies that the ratio of hot contents is almost constant regardless of date. Second, it is important to notice that most of traffic occurs from 5% video contents of the streaming server contents. TABLE presents more detail information in this connection. The statistics information in TABLE is about streaming server 1 at 01/01/2011 and shows the number of contents, total traffic, and traffic rate included its percentages.

In Table III, we can find 52.6% of the server's total traffic is consumed by 3% contents. Moreover, just 10% of the overall server contents occupy about three-quarter of the server traffic. This result shows the server traffic is so biased in a few hot contents while the other contents have miserly small traffic. It implies that by treating the hot contents properly, the streaming servers can effectively cope with high network traffic. There are some solutions about treating hot contents. One of the solutions is the uses of RAM Disk. By caching hot contents in fast RAM disk, the streaming servers can decrease the load of network traffic. Another solution is a fast identification of the hot contents. By distinguishing hot contents from cold contents, the streaming servers can distribute the hot contents to the other streaming servers in advance, and it can prevent the server from a sudden increase of the traffic.

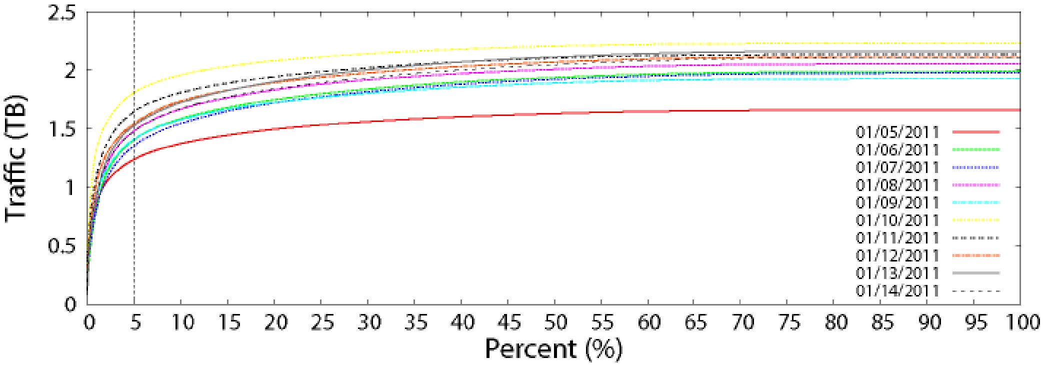


Fig. 4. CDF of Contents popularity distribution for streaming server 1

TABLE III: CONTENTS POPULARITY DISTRIBUTION OF STREAMING SERVER 1 (01/01/2011)

| | # of Contents | Traffic (MB) | Traffic (GB) | Traffic Rate (%) |
|---|---|---|---|---|
| 3% | 69 | 1,059,537 | 1,035 | 52.6 |
| 5% | 115 | 1,267,243 | 1,238 | 62.9 |
| 10% | 231 | 1,509,247 | 1,474 | 74.9 |
| Total | 2,351 | 2,016,179 | 1,969 | 100.0 |

## IV. RELATED WORK

There are many studies analysing traffic characterization about YouTube [2] VOD (Video on Demand), service. Micharel Zink et al observe the popularity of YouTube contents in a campus network and compared the observation with the global popularity [3]. Using DAG (Data Acquisition and Generation) card which captures packet headers are used, they collected traffic traces during 10-months and analyse it in various aspects like data rate of streaming sessions, the popularity of videos, and access patterns for video clips. From those observations, they confirm that trace has a comparatively stable statistics over a short-term period whereas long-term has a specific trend for a larger observation period.

The other study about YouTube traffic characterization is conducted by Phillipa Gill [4]. They collected traffic not the local (University of Calgary) campus network but also YouTube traffic for over a three month period. To collect traffic data, they use tcpdump [5] and used whois to determine the networks related YouTube. They find that the YouTube traffic similar with traditional Web and media streaming workload in terms of access patterns and video popularity characterization.

Rao at al.'s study on the network characteristics of Netflix and YouTube [6] show that the streaming strategies vary with the type of the application and the type of container used for video streaming. These strategies, three different streaming strategies with different traffic properties ranging from bulk TCP file transfer to non-ack clocked traffic, have a fundamental impact on the network traffic.

Analysis about P2P services also have occurred actively. Thomas Silverston at al introduces detail study about P2P IPTV community's traffic [7]. They collected the network traffic generated by 4 P2P IPTV applications. Using the traffic traces, this paper presents in-depth analysis about impact of the P2P application traffic on the networks, identification P2P application's traffic and P2P IPTV peer behaviour. There is, on the other hand, a study about Internet P2P streaming systems to observe traffic locality [8] and it shows PPLive [9] traffic is intensively localized at the ISP (Internet Service Provider) level.

Maier et al. study Mobile Hand-Held Device (MHD) and their impact on network usage using anonymized residential DSL broadband traces [10]. They analyse what kind of services users are interested in when they are at home and show that MHD traffic is dominated by multimedia content and downloads of mobile applications.

## V. CONCLUSION

We observe the traffic characteristics of N-Screen service in vary aspects by analysing the Apache logs generated from iShoot streaming servers. From the various approaches to the logs, we find some noticeable observations as following.

1) The daily traffic of a streaming server has a specific pattern. And it bound up with user life pattern.
2) The peak traffic time does not coincide with each

streaming server.

3) A few streaming servers had hot contents causing high network traffic, while the other servers have many cold contents accessed rarely.
4) 3% of a streaming server contents occupied above 50% of total traffic.

These characteristics have some significant implication which can contribute to design N-screen video streaming service. For example, it is important to identify hot contents as soon as it is uploaded, and if it is necessary, the contents have to be distributed to the several servers. In conclusion, our work represents a step towards understanding the features of N-Screen video streaming service.

## References

[1] IUMedia N-Screen Cloud Media Service Company. Available at http://www.ishoot.co.kr
[2] YouTube VOD service site Available at http://www.youtube.com
[3] M. Zink, K. Suh, Y. Gu, and J. Kurose, "Characteristics of youtube network traffic at a campus network-measurements, models, and implications," *Computer Networks* 53, 4 (2009), 501–514.
[4] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: A view from the edge," *In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (2007), ACM, pp. 15–28.
[5] Tcpdump, Available at: http://www.tcpdump.org
[6] D. Towsley, A. Rao, Y. Lim, C. Barakat, A. Legout, and W. Dabbous, Network characteristics of video streaming traffic, 2011.
[7] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, Pescap´E, A., G. Ventre, and K. Salamatian, Traffic analysis of peer-to-peer iptv communities. Computer Networks 53, 4 (2009), 470–484.
[8] Y. Liu, L. Guo, F. li, and S. Chen, "A case study of traffic locality in internet p2p live streaming systems," In Distributed Computing Systems, 2009. ICDCS'09. 29th IEEE International Conference on (2009), IEEE, pp. 423–432.
[9] PPLive, Available at:http://pplive.com
[10] G. Maier, F. Schneider, and A. Feldmann, "A first look at mobile hand-held device traffic," In Passive and Active Measurement (2010), Springer, pp. 161–170.