Report on Access Log Analysis Results for Streaming Delivery of the 2014 Summer Koshien

In live streaming of the National High School Baseball Championship held at Koshien Stadium (Summer Koshien) in August 2014, peak traffic of 108 Gbps was recorded, and there were approximately 1.9 billion requests. Here we examine the scale of access, as well as differences in access trends based on device types, as revealed by the results of analyzing the logs of all delivery servers.

3.1 Overview of Streaming Delivery of the 2014 Summer Koshien

IIJ provides streaming delivery services for Summer Koshien, which is produced by Asahi Broadcasting Corporation*1. Every year the special website set up by Asahi Broadcasting Corporation gets large numbers of hits, with most involving the receipt of live streamed content. Peak traffic of 108 Gbps was recorded during the final of the 2014 Summer Koshien on August 25, 2014.

In 2014, 38 Web servers were used for the live streaming of Summer Koshien (Figure 1). Asahi Broadcasting Corporation encodes the video, then uploads it to the ingest servers at IIJ using RTMP (Real Time Messaging Protocol). The ingest servers generate two types of content: HLS (HTTP Live Streaming) for mobile devices, and HDS (HTTP Dynamic Streaming) for PCs. These two types of content are cached on the web servers that act as the front end for clients, and the web servers return the requested content in response to requests from each client.

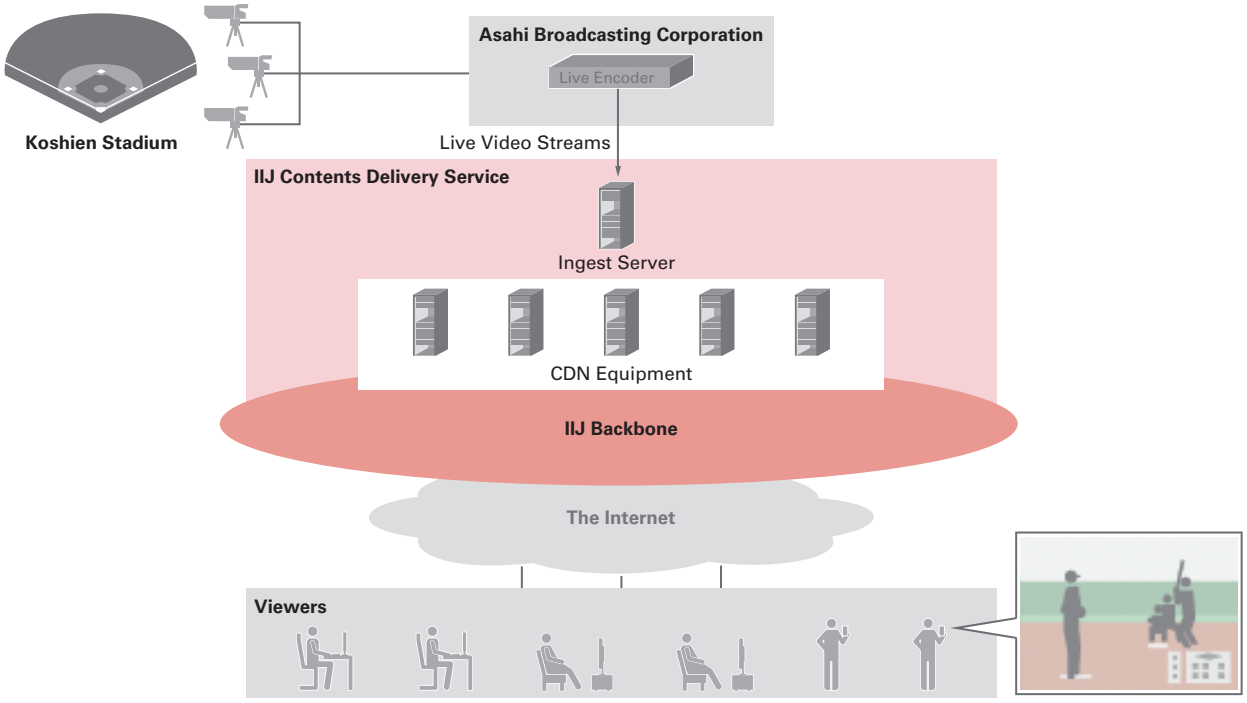

Figure 1: Conceptual Diagram for Delivery System

*1 Refer to Vol.25 of this report (http://www.iij.ad.jp/en/company/development/iir/pdf/iir_vol25_EN.pdf).

While watching a live stream, the client (a browser for PCs, or a dedicated application for mobile devices) downloads a playlist and segment files repeatedly. The playlist contains a list of segment files that can be downloaded at the current point in time, while the segment files contain video that is split into fixed intervals of time. Unlike the playback of long video files, clients must play back consecutive short segment files one after another to display a live stream. That means they must refer to a continually updated playlist and download segment files repeatedly to perform playback.

These repeated client requests for playlist and segment files are recorded along with the time of requests in access logs on each Web server. Because a large amount of access is concentrated in Summer Koshien live streaming, these access logs end up at the massive size of around 1.9 billion lines. The access logs contain a large number of records related to actual viewing activities of users, so by analyzing them we can understand the current state of user viewing trends and viewing quality in a production environment. We would like to utilize the knowledge we gained to improve the quality of streaming delivery in the future.

Table 1 indicates the scale of access for the live streaming of the 2014 Summer Koshien. The values shown here are calculated based on the access logs for all Web servers over the entire duration of the championship. The total number of requests matches the number of lines in the access logs mentioned earlier. The total amount of content sent is the total file size of each piece of playlist and segment file content recorded in the access logs. When content is sent from Web servers, headers for various protocols such as HTTP, TCP, and IP are also included. As a result, the volume of data sent from the Web servers actually exceeds the total amount of content sent.

There were 1.3 million unique IP addresses, and we found that a little over half of these (55%) were from mobile devices. This was the first time Asahi Broadcasting Corporation officially supported mobile devices by providing dedicated mobile applications for Summer Koshien live streaming, and it is now clear that many users actually viewed the live stream using mobile devices. When multiple clients receive live streaming via NAT, to Web servers it is seen as access from the same IP address, so in this case it is counted as a single unique IP address.

Table 1: Scale of Access for Live Streaming of the 2014 Summer Koshien

Total number of requests (hundred millions)	19.73
Total amount of content sent (TB)	531.4
Number of TCP connections (hundred millions)	2.81
Number of unique IP addresses (millions)	1.30

3.2 Changes in Access Numbers by Day and Hour

The 2014 Summer Koshien was postponed for two days due to rain, so it was held between August 11 (Mon) and August 25 (Mon), including a rest day on August 23 (Table 2). Usually, client numbers tend to increase as the championship progresses to the semifinals and final. To examine whether 2014 followed this pattern, we will first look at changes in access numbers during the championship based on the number of requests by day and hour.

Figure 2 shows changes in the number of hits by day. This bar graph indicates the number of hits per day. Before and after August 18, the trend in daily access numbers varies. Prior to August 18, the number of hits was lower overall, while for the following five days they were much higher. One plausible reason for the lower access numbers before August 18 is the Bon Festival that fell between August 13 and August 15. During this period many people took their summer vacation and were watching the Summer Koshien broadcast on TV at home, resulting in a decrease in live streaming views. The growth in the number of hits after the Bon Festival on August 18 and beyond is thought to be because people who were unable to watch the TV broadcast of the event during the week were watching via live streaming.

Next, Figure 3 shows changes in access numbers by hour. While Figure 2 showed the number of hits for each day, Figure 3 shows the number of hits per day divided into hourly increments. The number of requests on the day of the final is remarkable. The peak traffic was recorded on August 25 when the final was held, and at the same timing, the peak number of requests by hour was recorded. On the 25th the only game held was the final from 1:00 PM, and the high number of requests for the final is clearly demonstrated by the fact that request numbers by hour were more than double those for other game days.

Table 2: Schedule Overview for the 2014 Summer Koshien

Dates	Game Summary
August 11 (Mon) - August 14 (Thu)	Game 1
August 15 (Fri) - August 19 (Tue)	Game 2 (Game 1 for the first game on August 15 only)
August 20 (Wed) - August 21 (Thu)	Game 3
August 22 (Fri)	Quarterfinals
August 23 (Sat)	Rest day
August 24 (Sun)	Semifinals
August 25 (Mon)	Final

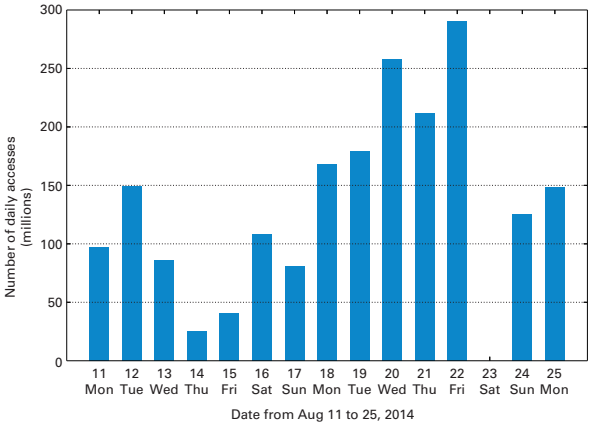

Figure 2: Changes in the Number of Hits by Day

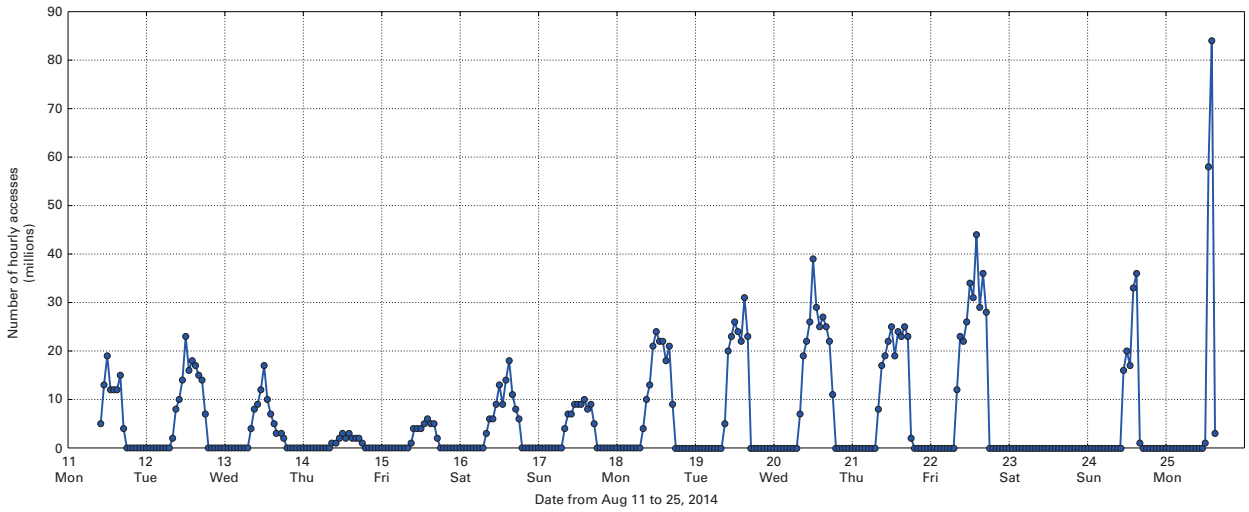

Figure 3: Changes in the Number of Hits by Hour

3.3 Differences in Viewing Activities by Device

Because this was the first trial of live streaming entire games for mobile devices, we will also examine differences between the viewing activities on PCs and mobile devices. Here we focus on the viewing time and viewing length.

3.3.1 Differences in Viewing Time

It is thought that viewing times vary for PCs and mobile devices such as smartphones because of differences in device usage. The first thing that comes to mind is that higher mobile device usage is expected during the times of day when people commute to work or school. Let us see if there were differences in viewing times based on device for the live streaming of Summer Koshien.

During the championship, there was some variance in the number of games held each day as well as the start time for games. If the numbers of games and start times are different, it is more likely that user viewing patterns will also change. For this reason, we selected eight days in which games were held from 8:00 am (August 13 - 14, 16 - 18, and 20 - 22), so we could target days where games took place around the same time of day. Of these eight days, six were weekdays, and the remaining two days fell on the weekend.

Using the access logs for these periods, we identified the number of requests by device for each hour between 8:00 am and 6:00 pm, and from this we calculated the proportion of access from mobile devices by hour. Figure 4 shows a box plot indicating the results of aggregating access ratios for each hour over the eight day period. The upper and lower parts of the blue boxes in the graph indicate the 75th percentile and 25th percentile, respectively. The red line in the boxes shows the median of the distribution. The median value for the 8:00 am timeslot was 71.8%, showing how high access from mobile devices was. The median value was 50% in the 9:00 am timeslot that followed, indicating that mobile devices accounted for half of the access numbers. After 10:00 am the median value fell below 40%, and continued fluctuating around 30%.

These results suggest that many users viewed the live streaming of Summer Koshien on mobile devices during the work and school commute period. After office hours began, users appeared to shift to viewing from PCs.

3.3.2 Differences in Viewing Length

The viewing time isn't the only aspect that differs between PCs and mobile devices. It is also conceivable that viewing length will vary. In the past, prolonged viewing of streaming video on the move was difficult for a number of reasons, such as the small screen size of mobile devices, and the lack of network bandwidth while mobile. However, there are now smartphones with large screens in addition to tablets, and the bandwidth available to mobile network environments has increased due to LTE and offloading to Wi-Fi. As a result, we are approaching the point where extended viewing from mobile devices is possible.

For comparison between the lengths of viewing on PCs and mobile devices, it is necessary to extract request sequences for each view from the access logs, and compare their lengths. However, user and viewing identifiers are not recorded in the access logs. Consequently, as an alternative we decided to shed light on trends in viewing length between PCs and mobile devices by comparing the length of consecutive segment file numbers.

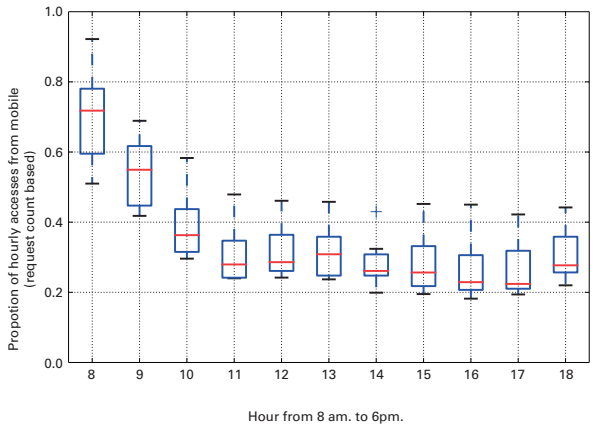

Figure 4: Hourly Mobile Device Access Ratios

Specifically, from access logs classified by client IP address and device type, we extracted only the segment file requests, then we identified places where there were five or more consecutive segment numbers in a request sequence. After this, we compared these numbers of consecutive segment files. We limited our research to places with five or more consecutive segment numbers, because it would not be appropriate to count places where the number of

segment files was too small to consider viewing. Due to the use of a method like this, we should make it clear that the number of segment files does not indicate the viewing time of each user.

Figure 5 shows the cumulative frequency distribution for the number of consecutive segment files on PCs and mobile devices. Comparing the number of consecutive segment files on PCs and mobile devices, those for PCs are around twice as long as those for mobile devices. Additionally, comparing median values we can see that those for PCs are 2.5 times longer than those for mobile devices. These results demonstrate that viewing length tends to be longer on PCs than on mobile devices. Furthermore, based on client-side measurement of viewing time, on PCs viewing time was about 20 minutes, while for mobile devices the viewing time was around 10 minutes. These results also point to viewing length being longer on PCs than on mobile devices.

3.4 Comparison of Client Numbers and Access Numbers by Device

In the previous section, we examined differences in viewing time and viewing length for PCs and mobile devices. Here we look into differences between PC and mobile device usage methods a bit further, based on the number of clients and hits.

Figure 6 shows daily changes in the number of clients for PCs and mobile devices, while Figure 7 shows daily changes in the number of hits from them. The two figures compare PCs and mobile devices day-to-day, but each indicates a different trend.

First, looking at the number of clients in Figure 6, we can see that the number of mobile device clients exceed the number of PC clients on almost all game days. Even on days where this wasn't the case, the number of mobile device clients was almost on par with the numbers for PC. Meanwhile, Figure 7 shows that the number of hits from PCs exceeded those from mobile devices on all game days.

As you can see, when we compare the number of clients and hits for PCs and mobile devices, mobile devices had high numbers in one case, and PCs in the other. This can be explained by the differences in the number of hits per client on PCs and mobile devices.

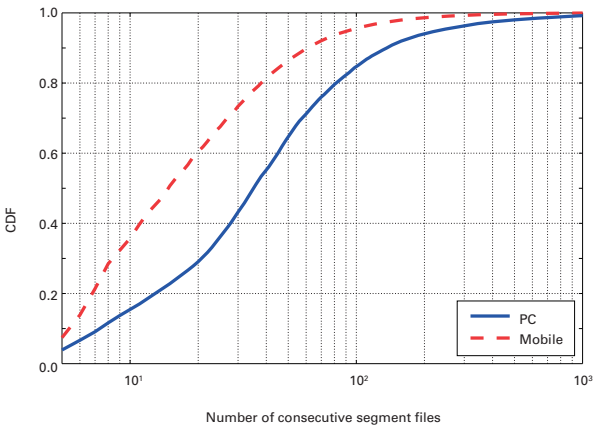

Figure 5: Number of Consecutive Segment Files Distribution by Device Type

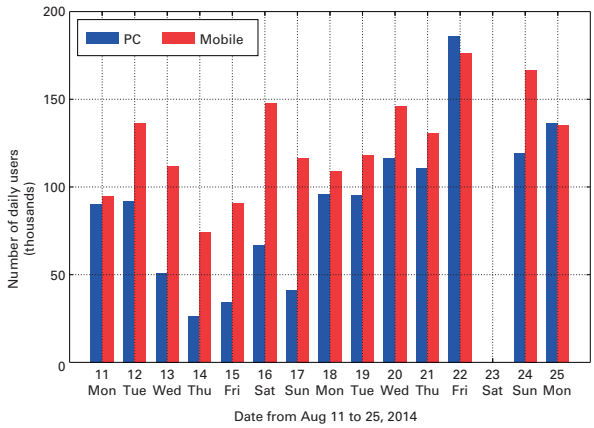

Figure 6: Comparison of Daily Client Numbers by Device

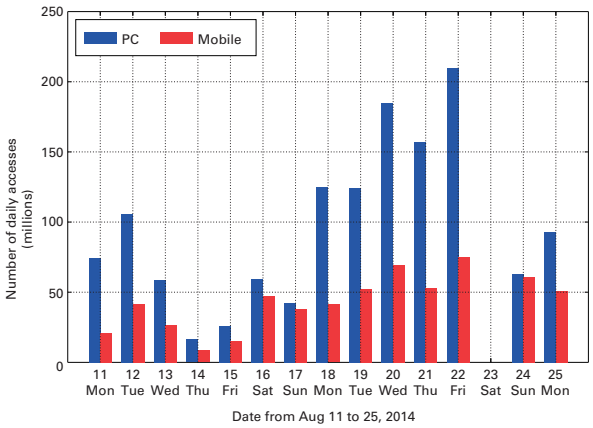

Figure 7: Comparison of Daily Access Numbers by Device

Figure 8 shows the cumulative frequency distribution for daily requests per client on PCs and mobile devices. Games took place for over 10 hours a day in longer cases, and if all 10 hours of streaming were viewed, a total of 4,500 segment files would be downloaded. Meanwhile, in Figure 8, PCs and mobile devices both converge on 1.0 around 5,000 where the green line is drawn. This suggests that although the ratio is small, some clients continued playback of video streaming for almost the full day.

Comparing the distribution of hits for PCs and mobile devices in Figure 8, we can see the access number distribution for each diverges from just after 10. The average for PCs is around 610 per client, while for mobile devices the average is around 158 per client, meaning access from mobile devices is only about a quarter that on PC. By understanding the extent of the difference in access numbers for each client on PCs and mobiles devices, we can comprehend that, even when the number of clients for mobile devices are higher than those for PCs in the abovementioned Figure 6, the number of hits for mobile devices are lower than those for PCs in Figure 7.

3.5 Conclusion

From the results of our investigation into live streaming for the 2014 Summer Koshien based on access logs for all delivery servers, we have seen changes in the scale and size of access, and identified differences in the access trends for PCs and mobile devices.

This was our first attempt at live streaming entire games to mobile devices via streaming delivery, and the analysis results revealed some differences in viewing trends between PCs and mobile devices. Examples include the fact that a particularly large number of viewers use mobile devices during the time when people commute to work and school, and the fact that mobile device viewers watch for shorter periods than PC viewers. Because the trend toward viewing streaming from mobile devices is set to accelerate in the future, we consider it crucial to analyze mobile device usage trends during the course of actual service to understand the viewing quality experienced.

Although we did not have the opportunity to discuss them this time, we are also conducting surveys and analysis based on the access logs used here with regard to viewing quality, including client behavior while viewing. We will continue performing surveys and analysis like this in the future, to help improve the quality of streaming delivery services.

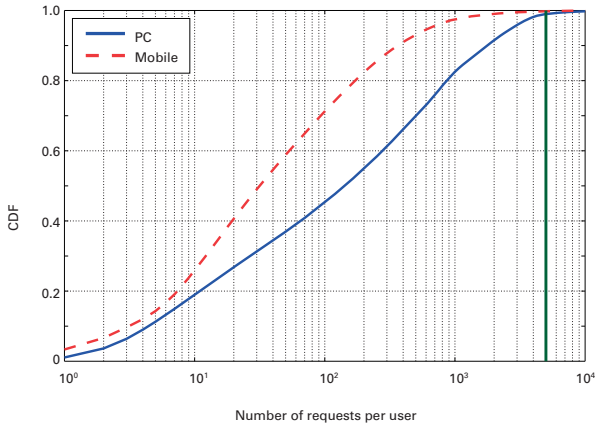

Figure 8: Access Number Distribution for Each Client by Device

Author:

Megumi Ninomiya
Ms. Ninomiya is a researcher at the Research Laboratory of the IIJ Innovation Institute. She is involved in research into Web traffic.