6 Forum Behavior Analysis

6.1 Background and Measurement Methodology

The Disney discussion forums at DISBoards.com are aimed at letting users share their Disney travel experiences and offer tips to others seeking a Disney vacation. According to Bigboards.com, the Disney discussion forums are the 28th largest forum site on the Internet with more than 34 million posts as of July, 2010 [60].

In a large forum site such as DISBoards.com, as shown in Figure 16, there usually exists a hierarchical structure. A forum is the smallest forum unit with one specific topic. In a forum, a thread denotes one discussion consisting of the original post by the thread creator and replies contributed by other users.

In order to quantify forum usage behavior, we randomly selected 21 Disney discussion forums and retrieved user and thread data. The 21 forums along with their thread and post counts are shown in Table 2. The crawl script went through every thread in the chosen forums that had received a post between 5/13/2010 and 6/13/2010; for each post on a thread, the post time and the user ID of the poster were collected. 13,807 threads were crawled in all. This yielded around 27,500 unique user IDs, representing 11.1% of DISBoards registered users. For every user ID, we fetch the statistics from the user’s DISBoards page, including the total number of posts and the forum name, date, and time of the 500 most recent posts. We were only able to collect the 500 most recent posts due to DISBoards search limits; however, only 4.8% of the users had more than 500 posts. Since we mainly focus on the behaviors of most users with the objective of improving the quality of experience of most users, incomplete sets of posts of 4.8% users will not greatly affect our observations of most users from the trace. To track views on a thread over time, we repeatedly crawled the forums for the desired time increment. Each of the 21 forums was crawled for views on 6/28/2010 and 6/29/2010 to determine views on threads for the one day period. Additionally, we tracked the views on threads in eight forums approximately every half hour from 6/23/2010 to 7/2/2010.

In establishing the design principles of MBoard, we follow the logic of: (1) Is there an increasing demand for multimedia sharing in forums? (2) Is P2P model suitable for multimedia sharing in forums? (3) What are the characteristics of forums including user activities and threads that we can take advantage of to optimize our design?

6.2 Is There an Increasing Demand for Multimedia Forums?

Our first question is whether there is an increasing demand for multimedia forums? An increasing demand for multimedia sharing will require forums to provide more bandwidth at a low cost, which necessitates the need of MBoard in forums nowadays. Figure 17 (right Y axis) shows the average size of posts including embedded media contents in bytes in a forum for each month from the period of Jan 2005 to Jan 2010. The figure shows a clear trend towards larger forum posts, which rises from 10kb to 80kb per post, supporting the notion that forums will be required to store and deliver more and more content. Additionally, Figure 17 (left X axis) also demonstrates that the trend is due to a greater number of multimedia elements including embedded 3rd party provided content in the posts; thus, providing users with an easier way to upload their content becomes a priority. Although the need for multimedia sharing is currently emerging in forums, most forums only provide very limited attachment size support, typically hundreds of KBs. This is insufficient to display high resolution images, so these multimedia files, including high-resolution photos and video clips, are linked from external 3rd party service providers. Storing and sharing the multimedia files among peers rather than the server can allow users to upload their large multimedia files directly to forums.

Figure 18 shows the estimated storage cost of different forum websites. We found the total number of posts on each forum as of July 26, 2010 from [60]. We multiplied the number of posts in each forum by the average post size (including the size of embedded media content) calculated from the trace data of DISBoards to get an estimate of the storage cost for each forum. The figure shows the storage cost of forums varies from 100GB to 10 TB. Thus, it is a challenge to the servers to serve such a tremendous amount of content to the vast number of users.

O1: The size of forum posts and the number of multimedia elements have been rapidly increasing in recent years.

What is the scale of MBoard deployment? Figure 19 shows the total number of users for each of the 21 DISBoards forums. It shows that the number of users in these forums ranges from 800 to 10,000, and the number is increasing yearly as shown in Figure 20. There could be even a larger number of anonymous users, so the number of users in forums may be very large. Figure 21 shows the number of threads in each forum. We can see that the thread count per forum varies widely from forum to forum. Although 60% of the forums have less than 5,000

<table>
<thead>
<tr>
<th>Forum</th>
<th>Threads</th>
<th>Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theme Park Attractions</td>
<td>234,799</td>
<td>3,134,586</td>
</tr>
<tr>
<td>Disney Resorts</td>
<td>183,318</td>
<td>2,114,282</td>
</tr>
<tr>
<td>Disney Restaurants</td>
<td>167,997</td>
<td>1,685,981</td>
</tr>
<tr>
<td>Camping at Disney World</td>
<td>114,744</td>
<td>252,074</td>
</tr>
<tr>
<td>Transportation</td>
<td>82,259</td>
<td>612,547</td>
</tr>
<tr>
<td>DVC-Operations</td>
<td>60,302</td>
<td>600,358</td>
</tr>
<tr>
<td>Budget Board</td>
<td>51,440</td>
<td>1,575,849</td>
</tr>
<tr>
<td>Disney For Families</td>
<td>51,440</td>
<td>738,883</td>
</tr>
<tr>
<td>DVC-Planning</td>
<td>38,073</td>
<td>339,610</td>
</tr>
<tr>
<td>Orlando Hotels &amp; Attractions</td>
<td>32,775</td>
<td>212,862</td>
</tr>
<tr>
<td>Teen Disney</td>
<td>31,341</td>
<td>1,419,451</td>
</tr>
<tr>
<td>DVC-Mousecellaneous</td>
<td>24,703</td>
<td>427,729</td>
</tr>
<tr>
<td>Disney Weddings</td>
<td>22,374</td>
<td>320,870</td>
</tr>
<tr>
<td>Disney Trip Reports</td>
<td>22,146</td>
<td>1,083,226</td>
</tr>
<tr>
<td>Disney World Tips</td>
<td>17,348</td>
<td>187,730</td>
</tr>
<tr>
<td>Welcome Board</td>
<td>15,858</td>
<td>162,995</td>
</tr>
<tr>
<td>DisABILITIES!</td>
<td>13,626</td>
<td>124,539</td>
</tr>
<tr>
<td>Disney for Adults</td>
<td>4,670</td>
<td>110,951</td>
</tr>
<tr>
<td>The College Board</td>
<td>3,730</td>
<td>66,812</td>
</tr>
<tr>
<td>Gay and Lesbian at Disney</td>
<td>3,526</td>
<td>89,112</td>
</tr>
<tr>
<td>Adventures by Disney</td>
<td>1,492</td>
<td>19,083</td>
</tr>
</tbody>
</table>

Figure 16: A forum structure.

TABLE 2: List of forums analyzed with their thread and post counts.
threads, the number of threads in popular forums is around 0.1-0.25 million threads. These figures indicate that the scale of a popular forum can be very large in terms of both the number of threads and users, which puts a tremendous strain on centralized servers, making their bandwidth a bottleneck.

O2: The number of users and threads in a forum can be very large, necessitating a scalable media sharing system.

Therefore, we should resort to a more efficient use of users’ resources. The P2P model is a promising method to tackle this problem. With this model, a peer retrieves its interested contents from other peers, and it resorts to the server if no other peer has the contents.

One question is how do we determine the suitable media size limit? Since there is no existing media sharing forum, we resort to YouTube statistics for the answer. Figure 22 shows the Cumulative Distribution Function (CDF) of YouTube video lengths. We see that 80% of the videos are approximately 300 seconds or less. Most of the remainder of the videos posted on YouTube range anywhere from (300,600] seconds. There are only a very small number of videos that have lengths exceeding 600 seconds. This is due to the fact that videos uploaded to YouTube cannot be longer than 10 minutes (i.e., 600 seconds) in length. Once, certain accounts were given the ability to upload videos exceeding 10 minutes in length, but this privilege is no longer given out [13]. Therefore, there are a small number of videos that exceed 600 seconds in length. We see that the large majority of videos posted on YouTube are short in length and that only 20% of all videos have a length greater than 300 seconds. In conclusion, most of the user generated contents are shorter than 10 minutes.

6.3 Is P2P Model Suitable for Multimedia Sharing in Forums?

P2P multimedia sharing requires that a certain number of peers exist in the forum to provide contents needed. A large number of users (O2) bring more benefits in using the P2P model, since more peers possessing a content file increases the content availability and downloading efficiency due to a more balanced load distribution. Figure 23 shows the bandwidth resource distribution from a MSN video trace collected by Microsoft, drawn based on the bandwidth of recorded users in an ascending manner. It shows that most users have decent bandwidth for video viewing with about 50% of users having more than 1Mbps and 35% of users having more than 2Mbps. This is important to the practicality of MBoard, since larger bandwidth brings faster multimedia download and sharing. By leveraging the bandwidths of all users in a forum, the load on the server can be greatly reduced.

The P2P model requires the existence of peers constantly viewing threads, so that the peers can help by uploading their watched content to others. Is there sufficient user online activity in a forum to facilitate P2P assistance? We focus on the characteristics of forum-level behaviors in order to prove that forum is suitable in peer-assisted environment. Figure 24 shows the activity time coverage of five randomly chosen popular threads in the Theme Park Forum, a medium sized forum out of the 21 forums. We see that in each time spot in the X axis, there are always some replies in the Y axis direction. It implies that there always exist users every minute in a forum. If we consider all threads in a forum, there would be more users existing in the forum every minute. If we combine these threads’ time coverage together, we see there are always some peers available that can be utilized to upload contents to others. This is the foundation for running multimedia forums in a P2P manner.

O3: There are always some users present in a forum. Moreover, popular threads receive constant views while unpopular threads receive few views in a day.

Then what are the characteristics of threads that MBoard can benefit from? Since all threads receive some views, we use replies to decide whether a thread has been active for some time. Figure 25 is a plot of the percentage of the total threads in each forum that never receive a reply. We see that an average of approximately 5% of the threads created in the forums never receive a reply. This tells us that approximately 5% of the total threads in a forum will fully depend on their server to keep the media content accessible due to the extremely unpopularity. But more importantly, this tells us that approximately 95% of the total threads in a forum will be active for some time, and
thus the media files in those threads can be stored in the P2P network for some period of time.

Figure 26 shows the total number of thread replies on the peak day. From this plot, we see that the average number of replies is somewhere around 30 and the highest number of replies is 120. This gives a good indication of the media content availability on the peak day. Once a thread hits its peak day, its media content will be widely available in the P2P network while it would otherwise exacerbate the load burden on the server in the server-client model.

Figure 27 is a plot of the CDF of the number of thread views over a 24 hour period in the 21 forums. As we see from the plot, 10% of the total threads have greater than or equal to approximately 300 views. 50% of threads have less than 30 views per day. The middle 40% of threads have medium-popularity with a view count between 30 and 300. We call the threads that receive a high number of views head content, and the ones receive a low number of views tail content. In conclusion, the amount of head content is small, but attracts many more views and results in higher P2P availability than tail content, while content with medium popularity can be made available by P2P to a certain extent.

Figure 28 shows the head/tail content contribution of seven randomly chosen forums. We define the traffic contribution factor as \( \frac{\text{total thread accesses}}{\text{number of threads}} \) and choose a cutoff of head and tail contents at the 20% of the total number of views of all threads. We can see that the head content on average has a contribution factor greater than 4 and contributes more than 80% of all thread viewing traffic, whereas tail content has a very small contribution factor and contributes less than 20% of the traffic.

O4: Most of the threads in the forums are tail content, while a small percentage of the threads in the forums are head content that contribute to most of the traffic, especially during the peak time.

6.4 What are the Characteristics of the Forums?

We would like to know whether a number of relatively stable nodes are always present in the forums, which can be exploited in the P2P model to enhance media content availability. We regard a stable node as a user that posts in one forum at least five times a day. The stable nodes are most likely the administrators and highly active users in the forums. This is an indirect measurement because collectable statistics do not provide the online time of each node.

From Figure 29, which was taken from a relatively popular forum, we see that the number of stable nodes is not constant over time. However, the average number of stable nodes is around 40, with a maximum of more than 100 and a minimum of 25. Figure 30 shows the number of stable nodes per forum. From this figure, we see that the number of stable nodes ranges from \([0,130]\) and 20 out of 21 forums have stable nodes.

O5: Certain nodes are almost always present in a forum; we call these stable nodes.

We consider a user to be online for 10 minutes if he posts/replies a thread. A user is considered to be continuously online for the duration if he continues to post at least once per hour. Figure 31 shows the average number of minutes spent online for 25,000 randomly chosen users. From this figure we see that most users spend an average of about 40 minutes online. Also, there are a small number of users that spend from 2 to 10 hours a day, which also confirms the existence of stable nodes. If we reduce the assumed online time of a user for one posing, we can still see from the figure that there exist relatively stable nodes though the number is reduced.

O6: According to our assumption that a user is online for 10 minutes if he posts/replies a thread or is continuously online if he keeps posting within 1 hour, users spend 40 minutes online a day on average, while some may spend many hours a day.

Another important question is the scope of nodes that MBoard should cluster in order to achieve the best communication efficiency. We should cluster nodes that always visit the same threads. Clustering nodes with less common interested threads will lead to unnecessary P2P structure maintenance cost and lookup cost in a large cluster. When a number of threads share many nodes that always reply and switch between these threads, we say these threads are highly connected by the user thread replying activities. A user may like to view different threads in one forum or switch
between different forums. We analyze the user thread replying activities to see whether some threads within a forum, all threads within a forum, or some threads in several different forums are more highly connected. Based on the observation results, we can know which group of nodes share more similar replying activities and hence should form into a P2P network in the design of MBoard.

Figure 32 shows the activities of 20 randomly chosen users in five forums. Each dot represents a different thread. A link between two or more threads indicates that a user posted a reply on one thread and then went on to the other thread(s) and posted a reply. By grouping the threads based on the forum, we observe that although some forums with more thread posting activities are connected closely, the lines are most densely connected within forums. There are connections between forums, but they are not as dense as those within forums. In the right cluster, the connections are sparse within the forum and it only has two connections with other forums. This is because the threads in this forum are not popular and few users visited the threads. This figure implies that the users switch between threads within a forum more frequently than between forums. Thus, we conclude that it is reasonable for MBoard to group threads by forum.

To confirm this conclusion, we study the connections between forums. Figure 33 shows the connectivity between the different forums. Each dot in this figure represents a different forum while each edge means a user switches from one forum to another. Here, we see that there are quite a few forums that are densely connected. However, there is also a fair number of forums that are not as densely connected. This figure confirms that the clustering feature of forums is not obvious and few users share interests in a number of the same forums. Thus, threads should be grouped by each forum rather than a group of forums.

The next question is whether we can further cluster all the threads within a forum. That is, whether the threads within a forum show a clustering feature. Thus, we study the connections between all threads within a forum. Figure 34 is a plot showing the connectivity of different threads visited by five randomly chosen users within a single forum. It is basically a map of the users’ posting activities within a forum. The high weight of the link connecting nodes is caused by two thread visiting patterns: (1) multiple users reply to the same thread and go on to reply to another thread in the same order, and (2) one user replies to a thread, proceeds to reply on another thread, and then returns to reply on the former thread. From this plot, we see that the majority of user activity is tightly clustered in one central location. Some thread reply paths tend to branch out in an area by themselves, but the number of these paths is relatively small when compared to the number of thread reply paths that are tightly connected. These threads are unpopular threads with fewer visits from users. This plot does not exhibit a clustering feature, which means that users in a forum tend to randomly visit all threads rather than preferring a certain group of threads. Therefore, all threads in a forum cannot be further clustered and forming all users in each forum into a P2P network is an optimal method.

Figure 35 shows the CDF of the number of different threads users post on. We see that 20% of users post on less than or equal to approximately 130 different threads. In fact, we see that only 50% of the users post on less than or equal to approximately 275 different threads, and 95% of users post on at least approximately 450 threads. This leads us to the conclusion that a large majority of the users tend to post on many different threads. Figure 36 shows the CDF of the number of different forums users post on. As we see from the plot, approximately 50% of users post on no more than 10 different forums. This increases to approximately 95% of users for no more than 25 different forums. We also see that only 20% of users post on less than approximately 5 different forums. From these observations, we see that the majority of the users tend to post on anywhere from 10 to 20 different forums. This is likely due to the fact that most forums are interest based and users tend to have a certain number of different interests. Therefore, the many threads posted by users are spread among only a few forums. This is a reason that we cannot find nodes sharing interests in several common forums (Figure 33). Thus, enabling nodes in one forum to share contents between each other is an optimal method. The observation results confirm that it is reasonable for MBoard to cluster the threads within one forum together.

Q7: Users in the same forum tend to view the same threads but tend to switch to different forums. That is, the thread viewing activities are clustered by forums.

In designing a system that uses a P2P structure where media content is pulled from other users who cached the content, we need to answer this question: what is the number of threads that users are interested in every day? Figure 37 shows the CDF of the number of different threads across forums that...
users reply to per day. Here we use reply activity to indicate strong interest in a thread. From this figure, we see that 90% of the users reply to approximately 3 different threads per day. Nearly 100% of the users reply to no more than 10 different threads per day. This indicates that most users are actually interested in a small number of threads in a day. Additionally, this implies that the number of pieces of media content cached in users’ computers could be small, so the cache burden will not be heavy for users.

Figure 38 shows that users seldom watch the same video again. From the distribution of users’ multiple playbacks of one video, we can see that most users do not watch videos they have seen before, meaning the cached videos are seldom used by users for replaying.

O8: Most users tend to reply to less than 10 threads per day, implying that most users are actually interested in a small number of threads. Therefore, they only need to have a small video cache.

## 7 Refreshing Scheme in MBoard

Based on O2, MBoard uses a refreshing time to discard content indices to content providers periodically to ensure the freshness of the indices and reduce communication cost. MBoard also does not need large cache according to O8.

In order to ensure the validity of the entries in the content table, upon departing, nodes should notify the brokers of the contents they are providing. A provider registered to a broker still may not be available due to a number of reasons: (1) it goes offline; (2) it stops providing uploading service; or (3) it deletes the cached videos. Therefore, the brokers need to update the index information in time in order to ensure that the chosen providers are in service. One way to deal with this problem is to let each node notify its broker before leaving. However, due to the high node join and departure frequency, this will generate a high communication overhead. In order to minimize the overhead, MBoard lets brokers automatically discard the registered indices which were reported a certain time period ago. We call this time period the refreshing time, denoted as \( r \). MBoard sets \( r \) equal to the continuous online time of the majority of the nodes.

We assume that each user can tolerate uploading its content for refreshing time \( r \). After \( r \), users can choose to continue to be in service if they are still online. MBoard can use tit-for-tat to assign more bandwidth to those users who upload more in order to encourage them to contribute their bandwidth for peer assistance. If a segment provider keeps sharing the segments after the refreshing time period, it will register with the broker again to refresh its service. Otherwise, the broker assumes this provider is no longer valid. The broker can assign each piece of the registered content a timer with a refreshing time value. When a piece of content is registered, the timer count-down begins. This piece of content will be automatically marked unavailable after \( r \). When this item is registered again, the timer will be refreshed. In this way, MBoard can ensure the availability of providers while reducing communication overhead.

O4 implies that a thread’s popularity deteriorates within days, so MBoard does not require users to hold a large cache such as 1GB (the typical cache size of the PPLive [6] client) for the forum content. The cache of each user is organized into a table for easy look up and service refreshing, called the cache table in MBoad. A typical user’s cache table is shown in Figure 39. In addition to the segmentID, the “count down timer” is set to the refreshing time. When it counts to zero and the user is still online, the service of the segment will be refreshed by re-registering. “Last used time” is used in knocking out the outdated cached items if the cache size limit is reached.

## 8 Additional Experimental Results

### 8.1 The Effect of Refreshing Scheme & Service Period

In testing the effect of the refreshing scheme, we assign users different continuous online time according to DISBoards’ user activities and our assumption in Figure 31 that if a user replies, its online time is 10 minutes and is increased if he replies again within 1 hour. Figure 40 shows the number of messages generated by service refreshing under different refreshing interval settings. We see that a small refreshing interval produces more messages than a large refreshing interval due to the more frequent communication between nodes and brokers. Figure 41 shows the P2P contribution percentage under different refreshing interval settings. We see that longer refreshing interval leads to lower P2P contribution. The reason that the server has to serve more requests is that most nodes have a short online time, so a long refreshing interval cannot let the broker update the availability information of segment providers in a timely manner. For refreshing intervals equal to 5 and 10 minutes, the P2P contribution remains very high because the majority of nodes have 10 minutes continuous online time. In conclusion, this experiment confirms our design that the optimal refreshing interval setting is equal to the online time of the majority of nodes in the network, which is 10 minutes in this experiment.

A node’s service period is the time period it is in service, i.e., the time period it can upload a video after starting to view a video. Figure 42 shows the CDF of the percentage of users versus the playback delay with different service periods. We see that the 3 minute service period leads to a slightly higher delay than other service periods. This is because when the service period is short, the server uploads more content and the queuing delay becomes larger. We also see that other service periods achieve a similar delay distribution. When the service period reaches a certain level, the nodes in service are sufficient to serve the new requests. In all cases, around 90% of users have delays no more than 0.8 seconds, which implies the effectiveness of P2P model for video sharing.
of unpopular videos, the server provides the service. This is due to the churn nature of forums where users constantly come and leave. A node needs to find video providers again if it has not received a response from a video provider who is leaving the system.

Figure 43 shows that the CDF of users’ playback interruptions varies slightly with different service periods, with the most interruptions occurring at a service period of 3 minutes. These results are caused by the same reasons as in Figure 42. In conclusion, the length of the service period does not have a great impact on delay or smoothness, and a certain service period decreases the video retrieval delay to a certain extent, with the optimal choice being 10 minutes.

Figure 44 shows the P2P contribution in MBoard over 7 days with different service periods. From this figure, we see that as the service period increases, the P2P contribution percentage increases as well. We see that if users are in service for only 3 minutes, MBoard reaches approximately 85% P2P contribution, while a 2 hour service period leads to approximately 90%. Overall, we see that the longer a user is in service, the larger the P2P contribution. Combining this observation with that of Figures 42 and 43, we conclude that MBoard should require users to be in service for at least 10 minutes (the online time of the majority of nodes in the network) as a contribution for a better user experience in the forum. This is easy to implement since a 10 minute service period is not much longer than the average length of a video.

### 8.2 The Effect of Cache Size

Figure 45 shows the CDF of the percentage of users versus the video playback delay with different cache sizes in each node. From this figure, we see that more than 95% of users have a delay that falls within the range of [0, 5] seconds, which is a well acceptable time for a user to wait. The rarely seen longest waiting time is around 35 seconds. Since larger cache size increases the video availability in peers, it is very intriguing to see that the effects of cache size on the video playback delay are almost negligible. Recall that users tend to view a small number of threads in a forum in O(5) and the number of viewers for popular videos is almost always constant. As a result, some providers for a request always exist even with a small cache size. If there is no provider for a request, such as for unpopular videos, the server provides the service. This result is mirrored in Figure 46, which shows that the CDF decreases as the cache size increases, with users’ playback interruptions barely changes with different cache sizes.

Figure 47 shows the effects of the cache size on the P2P contribution. From this figure, we see that the P2P contribution for a system with cache sizes of 60MB and 120MB are nearly identical. The system with cache size of 30MB has slightly lower P2P contribution. Combining this with the observations in Figures 45 and 46, we see that a larger cache size only improves the performance of the system slightly.

### 8.3 The P2P Contribution of Head/Tail Content

In Figure 48, top k% refers to the k percent of threads that have the largest number of accesses in our trace period, or the most popular threads. We see that the contribution factor (defined in Section 6.3) of head content is much higher than tail content. Moreover, 5% of the top threads (head content) have a contribution factor of 6, meaning they contribute up to 30% of all accesses, whereas 30% top threads contributes more than 70% of all accesses. Therefore, we conclude that popular threads have very high demand in multimedia forums, and they can benefit greatly from a P2P model. In addition, we know that most of the threads are tail content that generate a small amount of traffic. Since the absolute number of accesses for such unpopular threads is so small, they can be handled easily by the server.

We can see from Figure 49 that the P2P contribution is magnified more in smaller sets of popular threads, with the top 5% content having a contribution of around 98%, while unpopular threads have low P2P contribution. This is because the more popular a thread is, the higher possibility that a user will find other users watching the same thread. This shows that in MBoard, nearly all of the traffic of popular threads is handled by other peers instead of the server. This reduces the server load and the amount of server bandwidth consumed. When considering all traffic, the P2P contribution is around 90%, which also shows the overall P2P efficiency of MBoard remains very high. Additionally, we observe a very slight decreasing trend in the P2P contribution after it reaches its peak on day 3. This is due to the limited period of the trace file. At the end of the trace, there are a few new videos that are requested from the server, but these videos have few chances to be shared in our collected events.

Figure 50 shows a snapshot of the number of providers of each thread collected one hour before the entire experiment finishes when each node’s service period equals 10 minutes and 2 hours, respectively. The high results in the figure belong to the 2 hour service period test, which show that the number of available providers of these threads varies from 1 to 3. The number of available providers for the remaining threads in the 2 hour service period test is almost overlapped with all threads.
in the 10 minute service period test. Specifically, 3/4 of the results are 0 (left of the dotted red bar) and 1/4 of the results are 1 (right of the dotted red bar). This is consistent with our findings in O4 that the majority of threads are not very popular. The overlapping demonstrates that the number of providers does not differ much between the 10 minute and 2 hour service periods. This again supports our conclusion that MBoard can adopt the 10 minute uploading configuration. Several threads in the 2 hour service period test have more available providers, which is caused by its longer service period.

8.4 The Effect of Segment Size

As mentioned in Section 3.3, MBoard sets the segment size to 15 MBytes so that users do not have to split their videos and have lower probability to fail in uploading and downloading a large-size video. In the experiment for “segment size=15mb”, the size of videos is generated following the distribution of the embedded 3rd party video sizes in the trace data. In the experiment for “segment size=25mb”, we multiply the previously generated video sizes by a factor of $\frac{25}{15}$ in order to ensure that the size of majority videos is 25mb. Thus, the size of each video is also increased.

Figure 51 shows the CDF versus the video playback delay. We see that increasing the segment size increases the playback delay, especially for the users that already have a relatively high delay. This is because as the video size increases, users need more time to download the content, and the users waiting in the queue need to wait longer for them to finish downloading, leading to queuing time increase. Also, downloading larger size videos has higher probability to fail. From Figure 52, we see that the segment size influences the P2P contribution. As the segment size increases, the P2P contribution percentage decreases. This is because larger segments require more time to be downloaded and made available to other peers, which reduces the possibility for a peer to successfully locate a segment provider in the P2P network. We also observe that the P2P contribution percentage is low on the first day and remains nearly constant in the remaining days in all cases due to the same reason explained in Figure 9.