

EXHIBIT “A”

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF FLORIDA

CASE NO. 11-20427-WILLIAMS/TURNOFF

DISNEY ENTERPRISES, INC.,
TWENTIETH CENTURY FOX FILM CORPORATION,
UNIVERSAL CITY STUDIOS PRODUCTIONS LLLP,
COLUMBIA PICTURES INDUSTRIES, INC., and
WARNER BROS. ENTERTAINMENT INC.,

Plaintiffs,

v.

HOTFILE CORP., ANTON TITOV, and
DOES 1-10.

Defendants.

HOTFILE CORP.,

Counterclaimant,

v.

WARNER BROS. ENTERTAINMENT INC.,

Counterdefendant.

RULE 26(a)(2)(B) REPORT OF DR. RICHARD WATERMAN

1. My name is Richard Waterman and I am an Adjunct Professor of Statistics at The Wharton School at the University of Pennsylvania, and the President and Co-Founder of Analytic Business Services, Inc., a consultancy focused on providing expert advice and opinions in the field of statistical analysis. I received my Ph.D. in Statistics from the Pennsylvania State University in 1993. I have substantial experience designing and reviewing sampling protocols for various large organizations, such as the United States Postal Service, for whom I designed

and analyzed a national multi-stage sample for the estimation of operational characteristics. I have designed sampling protocols involving various filesharing technologies, specifically BitTorrent, Gnutella and Usenet. I also have substantial experience in designing sampling protocols in the private sector, and have developed market research studies for numerous large corporate clients, which typically involve issues related to sampling. Further details of my professional history, including a list of publications I have authored during the last ten years, can be found on the resume attached as Exhibit A. Within the last four years, I have testified as an expert at trial or deposition in the following cases, as further outlined in Exhibit B: *Arista Records LLC, et al. v. Lime Group LLC, et al.* No. 06-Civ. 05936 (S.D.N.Y.); *Columbia Pictures Industries, Inc. et al. v. Gary Fung*, No. 06-CV-5578 (C.D. Cal.); and *Schappell v. GEICO Corporation*, No. 1333 S2001 (Pa. Commw Ct.). I have submitted expert reports in *Columbia Pictures Industries, Inc. et al. v. Gary Fung*, No. 06-CV-5578 (C.D. Cal.); *Arista Records LLC, et al. v. Usenet.com, Inc.*, No. 07-CV-08822 (S.D.N.Y.); *Schappell v. GEICO Corporation*, No. 1333 S2001 (Pa. Commw. Ct.); *Freedom Medical Supply, Inc. V. PMA Capital Ins. Co.*, No. 003988 (Pa. Commw. Ct.); and *Blehm v. Albert Jacobs et al* 1:09-cv-02865-RPM, for the United States District Court of Colorado. I am being compensated for my services in this case at a rate of \$450/hour (\$550/hour for testimony).

2. I have been asked by the plaintiffs to create a protocol for drawing a statistically reliable sample for a study analyzing the percentage of files downloaded daily that were identified as infringing from the website operated by the defendants, www.hotfile.com (“Hotfile”).

3. I devised a methodology, described in more detail below, to allow for a scientifically reliable and unbiased sample of files to be selected from the population of interest.

After the sample had been drawn and the content obtained if available, an analysis of those files was conducted by a copyright analyst, Mr. Scott Zebrak, supervising a team in aid of his analysis. Mr. Zebrak's report describing the process he followed and his analysis is attached hereto as Exhibit C. For the determination of the copyright infringement status of each file in the sample, I relied on the work and conclusions of Mr. Zebrak.

4. Based on the analysis of the content files in the sample, I performed statistical analyses to derive the results for the infringement study. Those results are presented below, beginning with a summary of my opinions and conclusions and followed by a description of the study, the sampling protocol and analyses, and the bases and reasons for my opinions and conclusions. In general, in reaching my opinions and conclusions, I relied upon my specialized knowledge, education, and experience as applied to the facts and data discussed below, as well as data about downloads from Hotfile produced by defendants, and the work and conclusions of Mr. Zebrak. The exhibits I may use as a summary of or in support for my opinions are attached hereto or are being produced concurrently with this Report.

5. Based upon my review of the most recent data provided by Mr. Zebrak, approximately 90.3% of all daily downloads of files on Hotfile were downloads of infringing or highly likely infringing content; approximately 5.4% of the downloads of files per day on Hotfile were downloads of non-infringing or highly likely non-infringing files; and the remaining approximately 4.3% of the downloads of files per day on Hotfile were downloads of files whose copyright status could not be reliably determined in the time allowed. Of the works classified as non-infringing, 0.5% were identified as being likely illegal to distribute, making the infringement analysis here conservative. This analysis was based on data showing downloads of files that was provided by Hotfile.

6. The following describes the processes I used to design the sampling protocol and select the sample for the study:

7. The first step in devising the sampling protocol was to define the relevant population of interest from which the sample would be extracted, and to ensure the population was accurately represented in the sampling frame. Since the objective of the Hotfile study was to analyze the daily percentage of downloads of files from Hotfile that were of infringing files, the population of interest consists of downloads of files from Hotfile in a specified time prior to the complaint, January 2011.

8. While defendants did not produce actual log data for the period before February 2011, they did produce a data table called “dailydownload”, that efficiently summarizes all the necessary information that would be found in a log file to enable an infringement analysis of the recorded downloads. My understanding is that this table identifies files that were downloaded in a specific day (represented in the “uploadid” field), the date of download (represented in the “date” field), and the number of “premium” and “free” downloads of the files (represented in the “premium” and “free” fields). My understanding is that “premium” and “free” downloads are downloads by different kinds of users: those who have purchased Hotfile Premium subscriptions, and those that have not, respectively. Adding the two together gives the number of recorded downloads per day for the file on the indicated date. Thus, the “dailydownload” data contains a summary of information of recorded downloads by file for any particular day.

9. To understand the level of infringing activity on Hotfile prior to filing the complaint I looked at the month of activity prior to the complaint filing, January 2011. In order to understand the number of downloads per day in this month, I looked at different random days

in the month, and took a sample of downloads from each of those days. I designed the protocol to randomly select five weekdays and two weekend days.

10. In the first step of the protocol, I randomly selected five weekdays and two weekend days, by consecutively assigning each weekday in January 2011 a number and consecutively assigning each weekend day in January 2011 a number. I then used a standard random number generator to generate a separate list of numbers for the set of weekdays and the set of weekend days. This is a standard and universally accepted means to generate a simple random sample. The days selected by this process were January 5, 11, 20, 21, and 24 (weekdays) and January 1 and 30 (weekend days).

11. Overall, the “dailydownload” table shows 145,691,820 downloads of files from Hotfile in the month of January 2011. On each date selected, the “dailydownload” table shows the number of recorded downloads of files per day. The combination of the “free” and “premium” downloads per day for the selected days were as follows:

<u>Date</u>	<u>Download Count</u>
2011-Jan-01	4,180,329
2011-Jan-05	4,677,811
2011-Jan-11	4,568,087
2011-Jan-20	4,496,274
2011-Jan-21	4,631,944
2011-Jan-24	4,738,937
2011-Jan-30	5,125,537

12. Within each selected day, the sample frame was obtained by taking the dailydownload data and expanding the record of each file to capture the total number of recorded downloads of that file on that day. For example, if a file was downloaded 5 times in a day, the record would be expanded to reflect five separate downloads of that file. This method permits simple random sampling of the complete set of recorded downloads of all files in a day. The sample size was selected to obtain a 95% confidence interval with a margin of error of plus or minus 5%. (Because of the consistency of daily download infringement proportions, the final margin of error of the study was considerably smaller.) This allows for a high level of confidence that the results of the study reflect the percentage of infringing downloads per day for any day in the entire population, together with a high level of precision. To target this level of precision, I concluded that the Hotfile sample size should be 1750 (250 per day), which is also consistent with sample sizes in other similar online infringement studies conducted in other cases.

13. I used “simple random sampling” to draw the sample within each day. “Simple random sampling” is a universally accepted statistical methodology in which each item has the same opportunity to be chosen as any other item. In this case, each download of a file in a particular day had the same chance to be chosen as any other download of any file within that day. For each day, I used a standard random number generator to generate a list of numbers to select the downloads that constitute the sample. This too is a standard and universally accepted means to generate a simple random sample.

14. I am attaching herewith as Exhibit D the download instructions that implement the sampling protocol I have described in the foregoing for the Hotfile study. The protocol provides for replacement of files in the sample under only limited circumstances. First, if the file

appeared by its metadata to contain child or other illegal pornography, it was not included in the sample. Second, if the content file was corrupt, inoperable, or unplayable/undisplayable, for reasons other than being password-protected or encrypted, it was not included in the sample. In those cases, the files were replaced in the sample by another randomly selected file according to the protocol.

15. Mr. Zebrak provided an analysis showing his conclusions as to which of the 1750 sample files analyzed were determined to be either confirmed or highly likely copyright infringing, with the result broken down by download date. He also provided information as to which files he classified as highly likely or confirmed non-infringing, those “unknowable” files as to which no determination could be made, and “illegal” files that did not appear to be copyright infringing but that Mr. Zebrak concluded were likely illegal to distribute for other reasons. The infringement determinations of each download by day are itemized in the attached Exhibit E.


16. Based upon my review of the most recent data provided by Mr. Zebrak, by doing the calculations described above, I am able to conclude that approximately 90.3% of all daily downloads of files on Hotfile were downloads of infringing or highly likely infringing content; approximately 5.4% of the downloads of files per day on Hotfile were downloads of non-infringing or highly likely non-infringing files; and the remaining approximately 4.3% of the downloads of files per day on Hotfile were downloads of files whose copyright status could not be reliably determined in the time allowed. Of the works classified as non-infringing, 0.5% (nine files in the study) were identified as being likely illegal to distribute.

17. Using standard and universally accepted statistical methods to calculate a margin of error at a 95% confidence level yields a margin of error for this study of approximately 1.3%. This indicates a high level of reliability.

18. In my professional opinion, the sampling procedures used in the Hotfile study are based on standard and universally accepted statistical methods, and provide a scientifically valid sample from which we can reliably estimate the incidents of copyright infringement through the Hotfile website.

19. I continue to consider additional statistical analyses that might be conducted with additional data and/or time, including as to files that may have been uploaded to Hotfile but not downloaded, and reserve the right to supplement this report based on such further analyses. I further reserve the right to supplement or modify this report based on additional information that may come to light or based on further analyses.

Dated: November 18th, 2011


Richard Waterman, Ph.D.

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF FLORIDA

CASE NO. 11-20427-WILLIAMS/TURNOFF

DISNEY ENTERPRISES, INC.,
TWENTIETH CENTURY FOX FILM CORPORATION,
UNIVERSAL CITY STUDIOS PRODUCTIONS LLLP,
COLUMBIA PICTURES INDUSTRIES, INC., and
WARNER BROS. ENTERTAINMENT INC.,

Plaintiffs,

v.

HOTFILE CORP., ANTON TITOV, and
DOES 1-10.

Defendants.

_____ /

HOTFILE CORP.,

Counterclaimant,

v.

WARNER BROS. ENTERTAINMENT INC.,

Counterdefendant.

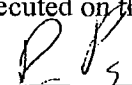
_____ /

CERTIFICATE OF SERVICE

I HEREBY CERTIFY on this 18th day of November, 2011, I served the following document on all counsel of record on the attached service list via the Court's CM/ECF filing system:

RULE 26(a)(2)(B) REPORT OF DR. RICHARD WATERMAN

I further certify that I am admitted to the United States Court for the Southern District of Florida and certify that this Certificate of Service was executed on this date.

By: 

Duane C. Pozza