IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

HARD DRIVE PRODUCTIONS, INC.

Plaintiff,

v.

DOES 1-1,495

Defendants.

Case No.: 1:11-cv-01741-JDB-JMF

Judge Bates

Magistrate Judge Facciola

DECLARATION OF SETH SCHOEN IN SUPPORT OF RECONSIDERATION OF DECEMBER 21 ORDER

I, Seth Schoen, declare as follows:

1. I am a Senior Staff Technologist with the Electronic Frontier Foundation (EFF), and I make this declaration on my own personal knowledge. I have worked with computers and computer networks for over a decade, have testified about electronic communications systems in two courts and before the United States Sentencing Commission, and have submitted declarations similar to my present declaration to the Federal courts in at least twelve other matters.

2. The purpose of this declaration is twofold. The first purpose of this declaration is to respond to assertions made by Plaintiff that might give a misleading impression of how unique BitTorrent is or how likely it is that various Defendants interacted with each other or were aware of each other in the course of uploading or downloading the motion picture whose copyright Plaintiff accuses them of infringing. The second is to set forth facts, which were readily available to Plaintiff from free, public Internet sources at and before the time it filed suit, that establish that many of the unnamed Defendants in the above-referenced case (hereinafter "Does" or "Doe Defendants") use Internet connections almost certainly located physically outside of the District of Columbia.

STATEMENTS RELATING TO MASS JOINDER

3. This Declaration responds to assertions made by the Plaintiff that might give a misleading impression of how unique BitTorrent is or how likely it is that various Defendants acted in concert in the course of uploading or downloading the motion picture whose copyright Plaintiff accuses them of infringing.

4. Plaintiff alleges that "unlike a traditional peer-to-peer network, each new file downloader is receiving a different piece of the data from each user who has already downloaded the file." Complaint ¶ 3. In fact, Defendants' behavior as alleged in the Complaint is strikingly similar to that of defendants who have used file sharing systems that were at issue in previous litigation about peer-to-peer file sharing, and to the extent it is different, the differences resulted in less direct communication among users of the technology, and less likelihood that any two

defendants worked in concert, not more.

5. First, BitTorrent is not the only system that has a swarming or multi-source download feature in which users can download simultaneously from several other users. Although this design was not a part of the earliest popular peer to peer systems such as Napster, it subsequently became quite widespread. For instance, the Kazaa and Gnutella software that was at issue in several copyright infringement actions have a swarming download feature that works similarly to BitTorrent's. *See, e.g.*, L. Jean Camp, "Peer to Peer Systems", in Hossein Bidgoli (ed.), *The Internet Encyclopedia* (Wiley, 2004), vol. 3, at 30. ("In order to increase the speed of downloads and distribute the load on peer-provid[ed] files Limewire uses swarming transfers. Swarm downloading entails downloading different elements of files available on multiple low-bandwidth connections to obtain the equivalent service of a single broadband connection."); *see also* Alex Jantunen *et al.*, "Peer to Peer Analysis: State of the Art" (Tampere University of Technology, 2006) (noting that swarming supporting protocols include at least FastTrack, Gnutella, ED2K/Overnet and BitTorrent).

6. Second, BitTorrent's file-focused distribution provides users with *less* ability to identify and communicate with the peers with whom they exchange files than other technologies do. For example, Napster and KaZaA, unlike BitTorrent, referred to each user by a human-intelligible and somewhat memorable screen name, instead of a number. Napster and KaZaA have also offered users the ability to chat with one another. BitTorrent does not offer these features. There is no easy way for the various BitTorrent users who have uploaded or downloaded parts of a file to recognize, name, or communicate with one another.

7. While BitTorrent client software, like other peer-to-peer file sharing software, may provide a way for a user to view the IP addresses of peers, users are not required to do so in order to use BitTorrent. They do not have to select peers' IP addresses, because the selection of peers is done automatically. Indeed, since BitTorrent automates so much of the download process, many users likely do not even know how BitTorrent works. Most BitTorrent users have no reason to know how many or which other peers they might have communicated with in the

course of downloading a file, or which addresses transmitted which portions of the file.

8. For example, the main screen of the popular Azureus BitTorrent software shows only a progress bar for the download, indication the percentage of the download that is complete, without mentioning other any other peers or their Internet addresses. Although interested users can learn about the role of peers or view their IP addresses, they are not required to do this.

9. I do not believe Plaintiff could have obtained direct evidence that any particular defendant shared portions of the copyrighted work at issue here with any particular other defendant, since BitTorrent does not provide a means for third parties to learn directly who is downloading files from whom.

10. Moreover, the plausibility that a given user downloaded a part of a file from any other particular user rapidly evaporates as the number of users becomes larger or as the users use BitTorrent at widely separated times. Both are true in this case. The number of users sued together in this case is 1,495 and, according to the records submitted by Plaintiff, they allegedly used BitTorrent at different times over nearly four months.

11. Both of these facts — the number of individuals named together and the different times of their alleged use of BitTorrent — make it highly implausible that all of 1,495 individuals sued jointly here uploaded or downloaded a part of the file from each other.

12. As to the different times for download specifically, the various Defendants are alleged to have used BitTorrent to transfer the movie file at very different times over the course of 117 days, which makes it even less plausible that they all could have communicated with one another. Exhibit A to the Complaint shows allegations of infringement on dates ranging from May 29, 2011, through September 23, 2011. Consistent with academic research on file-sharing using BitTorrent described below, this shows another reason why many individual defendants would never have communicated with one another: although some BitTorrent users may continue to share a file for a period of time after their download has completed, most do not.

13. Empirical research shows that most BitTorrent users do not remain connected for very long after their downloads are complete. These statistics can be measured by means quite

similar to the techniques employed by Plaintiff here. One large study observed that only 3.1% of BitTorrent users stayed connected (to upload to others) more than ten hours after their downloads completed; only 0.34% stayed connected over 100 hours. J. A. Pouwelse, P. Garbacki, D. H. J. Epema, and H. J. Sips, *The BitTorrent P2P File-Sharing System: Measurement and Analysis* at 4, in Proceedings of the 4th International Workshop on Peer-to-Peer Systems, *available at* http:// iptps05.cs.cornell.edu/PDFs/CameraReady_202.pdf.

14. Another study found that over 90% of users who successfully downloaded a file remained connected for less than a single day, while many users who attempted to download the file gave up entirely and disconnected within the first few hours. M. Izal, G. Urvoy-Keller, E. W. Biersack, P. A. Felber, A. Al Hamra, and L. Garcés-Erice, *Dissecting BitTorrent: Five Months in a Torrent's Lifetime* at 7, in Proceedings of the 5th International Workshop on Passive and Active Network Management Proceedings of the 4th International Workshop on Peer-to-Peer Systems, *available at* http://www.pam2004.org/papers/148.pdf.

15. Thus, it is highly unlikely all or even a significant number of the defendants who downloaded the subject copyrighted work here stayed on the network and became a source for another later-connecting defendant to download from days or weeks later.

16. In addition, according to Exhibit A of the Complaint, the Defendants allegedly participated used BitTorrent to infringe Plaintiff's work sometime between May 29, 2011 and September 23, 2011. This leads me to believe that it is very unlikely that any of these defendants directly communicated with more than a few others. Research (cited in paragraphs 13 and 14 above) shows that most BitTorrent users only remain in swarms for quite short times after finishing their downloads, usually for less than a day, not nearly for the 117 days between May 29 and August 23.

17. Plaintiff alleges that 1,495 defendants were "simultaneously stealing copyrighted material from many ISPs in numerous jurisdictions around the country," apparently basing this allegation on the way that BitTorrent works. Complaint ¶ 4. This statement could create the misconception that each participant in a swarm communicated or exchanged portions of a file

with every other participant. In fact, a downloader receives a given segment of the file from only one other user, not from all of the users participating in a given swarm. BitTorrent does not permit downloading a particular piece of a file from more than one user at a time, although different pieces of the file can be downloaded from different users. Also, a downloader only communicates with some of the users in a limited, gradually changing "peer set" of generally no more than 50 peers at a time. While it is quite plausible that some Doe Defendants shared some pieces of the allegedly infringing file with some of the other Defendants, it is far from clear (and even unlikely) that all of the Defendants worked in concert in downloading a single file in the sense of communicating with one another or exchanging portions of a file with one another.

STATEMENTS RELATING TO PERSONAL JURISDICTION

18. By reviewing Exhibit A to the Complaint, I compiled a list of the Internet Protocol (IP) addresses that Plaintiff attributes to each of the Doe Defendants.

19. There are many tools freely available to the public that help reveal where a person using a particular IP address is likely to be physically located. This process is often referred to as "geolocation." This information is commonly used for many purposes, such as customizing the language or content of web sites based on inferences about where visitors are accessing the site from. For example, Google, Inc., uses geolocation to choose to display its web site in German to people coming from Germany, in French to people coming from France, and so on. It also uses geolocation to display ads and results related to particular cities or regions to people accessing its site from those cities or regions.

20. One means of learning about where an IP address is physically located is known as "reverse domain name service lookup" or "reverse DNS." When an Internet service provider ("ISP") allocates or prepares to allocate IP addresses to customers, it typically creates and publishes database records assigning a human-readable "domain name" to each numerical IP address. The reverse lookup information can be obtained by anyone using a program such as "host," which is a standard program included with many computer operating systems, or with any of several web-based tools such as the DNS lookup service at http://lookupserver.com/.

21. One of the purposes of reverse DNS is to help interested parties learn more about what a computer is used for, what organization's network it is connected to, and, in many cases, where the computer is physically located. Typically, for home users of dial-up or broadband connections, such as DSL or cable-modem services, a domain name obtained from reverse DNS will identify which ISP assigned the IP address.

22. In addition, such a domain name will frequently incorporate an approximate physical location, such as the name of a municipal area, state, or region. For example, Doe #28 is identified by the IP address 108.2.97.190 and is described by Plaintiff as a subscriber of Verizon Online. The reverse DNS database identifies this computer as pool-108-2-97-190.phlapa.fios.verizon.net, confirming Plaintiff's suggestion that Doe #28 is a Verizon customer, but adding the additional detail that the likely physical location of the computer is in or near Philadelphia, PA ("phlapa"). This means that in all likelihood, the individual who used this IP address is located in Pennsylvania.

23. Although Internet service providers are not required to publish this information, and although it is sometimes only given to state-level precision, it can, when available, be a useful source of data about where an individual Internet connection is most likely located.

24. For each of the 1,495 IP addresses that were references in this suit, I used the "host" program to perform a reverse lookup against the publicly-accessible reverse DNS service.

25. The results of this process generally confirmed Plaintiff's association of particular IP addresses with particular ISPs. Additionally, the results of this process generally strongly suggested a geographic location for most individual defendants. In other words, most of the Does listed in this lawsuit can be associated by the host reverse DNS look-up with both an Internet service provider and a geographic location.

26. Reverse DNS records indicate that Does in this lawsuit include customers with Internet connections located in virtually all areas of the United States, including some in or near San Francisco, Minnesota, Pennsylvania, Georgia, Texas, Florida, New York, and other states and regions throughout the United States.

27. In addition to reverse DNS information, another means of learning where an IP address is located is to use "geolocation databases." Several companies collect and continually update geographic information about IP address locations from a variety of data sources, and collect this information in databases called "geolocation databases." Geolocation databases are commonly used by web site operators who are interested in finding out the approximate physical location of their web visitors. Since web site operators are often very interested in such information, there is considerable demand for geolocation databases.

28. Geolocation databases may be sold or given away for free. One very popular geolocation database is the "GeoIP" database maintained by MaxMind, Inc., a Boston company that specializes in geolocation technology. In addition to other sources of information, MaxMind explains that it "employ[s] user-entered location data from sites that ask web visitors to provide their geographic location" in order to learn which IP address ranges correspond to which cities and states. MaxMind, http://www.maxmind.com/app/ip-locate.

29. A version of the MaxMind GeoIP geolocation database is freely available for anyone to download from MaxMind. The company claims that this free version can determine the location of "79% [of U.S. IP addresses] within a 25 mile radius." MaxMind, http://www.maxmind.com/app/geolitecity.

30. I downloaded this freely available database and looked up each mentioned IP address in it, obtaining an estimated city and state location for each such address.

31. Because DSL and cable modem connections are provided from local hubs to users in a particular geographic region, there is good reason to believe that the geographic location data obtained by these methods actually reflects the physical location of the Internet connection, at least in general terms. In other words, although geolocation data is not perfectly accurate, the geographic designations obtained by these methods likely indicate the approximate locations of the residences or other venues where the Does use their Internet-connected computers.

32. I have attached hereto as Exhibit A to this Declaration a list of the reverse DNS names of the Doe Defendants' distinct IP addresses, as well as the estimated physical location of

each such IP address according to the freely available version of the MaxMind GeoLite City database.

33. Though the MaxMind GeoLite City database and reverse DNS records are not perfectly accurate, I know of no reason to think that either source of information has a bias that makes it more or less likely that an individual IP address will appear to be located in the District of Columbia.

34. From the information available from the MaxMind geolocation database, 8 of the IP addresses appear to be located in the District of Columbia, 1,479 outside of the District, and 8 are not assigned to any location by the database. This puts around 0.5% of the IP addresses in the District of Columbia, comparable to the 0.2% of the population of the United States as a whole that resides in D.C. according to the 2010 Census.

STATEMENTS RELATING TO INTERNET USE GENERALLY

35. Most Internet users in the United States use residential broadband services that they, or someone in their household, contracted for with a commercial Internet service provider. As a part of contracting for these services, the subscribers generally identified themselves to the providers.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge, and that this declaration was executed in San Francisco, California.

Dated: January 26, 2012

By: ______________________
SETH SCHOEN