Corynne McSherry, Esq. (SBN 221504)
Matthew Zimmerman, Esq. (SBN 212423)
ELECTRONIC FRONTIER FOUNDATION
454 Shotwell Street
San Francisco, CA 94110
Telephone: (415) 436-9333
Facsimile: (415) 436-9993
Email: corynne@eff.org

Attorneys for *Amicus Curiae*
Electronic Frontier Foundation

UNITED STATES DISTRICT COURT

CENTRAL DISTRICT OF CALIFORNIA

CAMELOT DISTRIBUTION GROUP, INC., Plaintiff, v. DOES 1 THROUGH 5865 Defendants.	Case No. 11-cv-01949 DDP (FMOx) **DECLARATION OF SETH SCHOEN**

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 seven other matters.

2. The purpose of this declaration is twofold. The first purpose 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 physically located outside of the State of California. The second 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.

STATEMENTS RELATING TO PERSONAL JURISDICTION

3. By reviewing Exhibit B to the Declaration of Scott Plamondon ("Plamondon Decl."), I compiled a list of the Internet Protocol (IP) addresses that Plaintiff attributes to each of the Doe Defendants.

4. Sometimes the same Internet Protocol address is used by more than one of the Doe Defendants. For example, consider 67.185.165.231. This IP address is being sued as Comcast Doe #1086 (*id.* at docketed page 115), Comcast Doe #1121 (*id.* at docketed page 117), and Comcast Doe #76 (Plamondon Decl. at docketed page 325). Plaintiffs may have named the same IP address multiple times because they observed the same IP address participating in BitTorrent transfers at different times, and it is possible that the IP address was being used by a different human Internet subscriber each time.

5. I found Plaintiff's numbering scheme for Doe Defendants confusing in comparison to the numbering schemes used by plaintiffs in other copyright litigation in which similar numbers of defendants were sued. Plaintiff here, rather than using a single consistent number for each defendant, has restarted the

defendant numbering for each and every Internet service provider; for example, there are Doe Defendants #1, #2, and #3 from Bellsouth.net (Plamondon Decl. at docketed page 13), other Doe Defendants #1, #2, and #3 from Cellco Partners (*id.* at docketed page 29), still other Doe Defendants #1, #2, and #3 from CenturyTel (*id.* at docketed page 31), and so on for each individual ISP whose subscribers are being sued. What's more, some ISPs appear more than once in the list; Comcast Cable subscribers are listed beginning at docketed page 52 and again beginning at docketed page 321 – yet different Comcast subscribers are given the same Doe Defendant numbers in the two lists!

6. Because I see no concise and straightforward way to refer to individual Doe Defendants and because my present declaration concerns the distinct IP addresses mentioned by Plaintiff rather than the Doe Defendants, I used software to create a list of all of the distinct IP addresses from Exhibit B to the Plamondon Decl. This process identified 5399 distinct IP addresses, which I have chosen to list in ascending numerical order for ease of reference. The numerically least IP address is 24.0.116.212 (which I'll refer to as "IP address #1") and the numerically greatest is 216.40.145.226 (which I'll refer to as "IP address #5399").

7. 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.

8. 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/>.

9. 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.

10. 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, one of the Does being sued here — Comcast Cable Doe Defendant #1001, mentioned on page 110 of Exhibit B to Plamondon Decl. — is identified by the IP address 98.195.59.238 and described by Plaintiff as a subscriber of Comcast Cable. (According to my numbering of the IP addresses, this address is IP address #4121.) The reverse DNS database identifies this computer as c-98-195-59-238.hsd1.tx.comcast.net, confirming Plaintiff's suggestion that Doe #1001 is a Comcast ("comcast.net") customer, likely using Comcast's cable Internet service, but adding the additional detail that the likely physical location of the computer is in or near Texas ("tx"). This means that in all likelihood, the individual who used this IP address is located in the State of Texas.

11. 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.

12. For each of the 5399 IP address that were referenced in this suit, I used the "host" program to perform a reverse lookup against the publicly-accessible reverse DNS service.

13. 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.

14. 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 Michigan; Massachusetts; New York City; Tampa Bay, Florida; Hawai'i; Maryland; New Jersey; Washington; and other states and regions throughout the United States.

15. In addition to reverse DNS information, another means of learning where an IP address is located is to use a public database operated by the American Registry for Internet Numbers ("ARIN"). ARIN is the authority responsible for the initial allocation of IP addresses to ISPs located in the United States. ARIN maintains public records indicating to whom a given IP address has been allocated. Large ISPs may apply to ARIN multiple times to receive multiple "blocks" or ranges of IP addresses. Each such block may be dedicated to a particular purpose or geographic area.

16. The ARIN database can be searched using a public web site provided by ARIN at https://www.arin.net, or by using a program called "whois," which is a

standard part of some operating systems and performs the same database-searching function. There is no charge for searching the ARIN database.

17. For example, Doe Defendant #268 on page 187 is identified by the IP address 173.168.125.85. I searched the ARIN whois database for this address and learned that this address is part of a network assigned to "Road Runner HoldCo LLC." The whois record also contains a comment asserting that the network "serve[s] Road Runner residential customers out of [...] Austin, TX and Tampa Bay, FL." This information is readily available at <http://whois.arin.net/rest/org/RRSW>. Combined with the reverse DNS record for this IP address, which is cpe-173-168-125-85.tampabay.res.rr.com, there is a strong inference that the user of this IP address resides in or around Tampa Bay, Florida.

18. In addition, 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.

19. 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> (last visited May 17, 2011).

20. 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> (last visited May 27, 2011).

21. 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.

22. 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.

23. 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.

24. In my experience, computer professionals are generally aware of the existence and function of the reverse DNS and whois services, as well as geolocation databases such as the GeoIP database, and would use any or all of these sources of information when they needed to learn where a given IP address was physically located. These techniques are readily and easily available to Plaintiffs, their attorney, and to the computer professionals they have employed to perform the investigations leading to this lawsuit.

25. 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 California.

26. From the information available from the MaxMind geolocation database, 734 (seven hundred thirty-four) of the IP addresses appear to be located in the State of California, 4606 outside of California, and 59 are not assigned to any location by the database. This puts around 13.6% of the IP addresses in the State of California, compared with the 12.1% of the population of the United States as a whole that resides in California according to the 2010 Census.

27. Separately from the question of where Does reside, Plaintiffs did not submit all the details of the investigations that led them to accuse these Does of copyright infringement. These details could be important because simple methods of attempting to locate copyright infringers can easily go awry. For example, in 2008 researchers from the University of Washington found that, given then-prevalent methods for investigating BitTorrent transfers, it was straightforward to frame particular IP addresses for downloading files that they had not, in fact, ever attempted to download. The researchers experimentally framed their own laser printer and succeeded in eliciting false allegations of copyright infringement against it. *See* Michael Piatek, Tadayoshi Kohno, and Arvind Krishnamurthy, "Challenges and Directions for Monitoring P2P File Sharing Networks, or, Why My Printer Received a DMCA Takedown Notice," in *Proceedings of the 3rd USENIX Workshop on Hot Topics in Security*, July 29, 2008, *available at* http://www.usenix.org/event/hotsec08/tech/full_papers/piatek/piatek.pdf.

STATEMENTS RELATING TO MASS JOINDER

28. I reviewed the Declaration of Tobias Fieser in Support of Plaintiff's Response to Order to Show Cause ("Fieser Decl."), as well as Plaintiff's Response to Order to Show Cause ("Pl.'s Resp."). I also reviewed some of the academic research on BitTorrent, as indicated below.

29. 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 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.

30. Plaintiff claims that BitTorrent is "significantly different in form from the older P2P protocols . . . such as Napster, Kazaa, Limewire, and Gnutella." Fieser Decl. ¶ 2. In support of this claim, Plaintiff points to two specific aspects: the nature of BitTorrent's "swarm downloads," Fieser Decl. ¶¶ 4-7, and its "file-focused" — as opposed to "user-focused" — method of file-sharing. Fieser Decl. ¶ 9.

31. However, BitTorrent is actually strikingly similar in one important regard to 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 result in less direct communication among users of the technology, not more.

32. 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).

33. 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.

34. 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.

35. 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. *See, e.g.*, <http://torrent-search.us/images/torrent-clients/azureus-screenshot.jpg> (screenshot of Azureus software in the midst of a download). Although interested users can learn about the role of peers or view their IP addresses, they are not required to do this.

36. I do not believe Plaintiff's experts 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.

37. 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 in over five thousand and, according to the records submitted by Plaintiff, they allegedly used BitTorrent at different times over the course of two months**.**

38. 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 the 5,685 individuals sued jointly here uploaded or downloaded a part of the file from each other.

39. 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 two months, which makes it even less plausible that they all could have communicated with one another. Appendix B to Plamondon Decl. shows allegations of infringement on dates ranging from January 11, 2011 through March 1, 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 other defendants: although some BitTorrent users may continue to share a file for a period of time after their download has completed, most do not.

40. 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's experts 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://www.springerlink.com/content/l251rj12233u05l>.

41. 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.springerlink.com/content/fg8hqw4136t0vtx9/>.

42. 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.

43. Plaintiff states that "each downloader is receiving a different piece of the data from users who have already downloaded that piece of data." Pl.'s Resp. 4. This statement could create two misconceptions about how BitTorrent works. In fact, a downloader receives a given "piece" of the file from only one other user, not from all "users who have already downloaded that piece." 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. While it is possible that *some* Doe Defendants shared *some* pieces of the allegedly infringing file with *some* of the other Defendants, Plaintiff's assertion that "all of the Doe Defendants acted in concert with one another" with the others is unsupported by its factual allegations or the nature of the BitTorrent protocol. *Id*. at 7.

44. Plaintiff also states that "all of the events involving all of the Doe Defendants are logically related" to the original infringer's decision to start sharing a particular version of a motion picture. *Id*. at 6. This assertion is flatly contradicted by Plaintiff's own evidence. On the first page of Exhibit B, Plaintiff mentions seven different versions of the allegedly infringing file, based on the "File Size" and the "File Hash." Plamondon Decl., Ex. B. Thus, there are at least seven — and in fact, many, many more — original infringers. The Plamondon Declaration ultimately mentions twenty different hash values (namely 2M6OSD, 3IOKAC, 77VY6, 7MNIJ2, AILL4P, BBFKS, BWSN, IZJ3L5, JKTXT, JP76VD, LLZXB, OLSHND, PG3WM, U3B44, URL6A, VQDVD, WTASE, XWPJN, XXGCV, and ZKLTB). Each of these refers to a separate and independent copy of the motion picture.

45. While a single Defendant may have a "logical relationship" to the original infringer of his version of the file, he has in fact no relationship to the original infringer of other versions of that file.

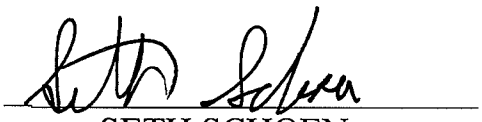
46. As I stated earlier, many other modern P2P systems do support swarming downloads akin to BitTorrent's, so it is hard to be confident that an infringer basically copied a work from one other user in the incidents at issue in prior file sharing litigation.

47. In any case, in all peer-to-peer file sharing networks, particular files can become more widespread throughout the network over time as new users obtain them from earlier users. Indeed, researchers have been able to quantify and analyze

the spread of particular files in particular networks over time. Regardless of whether particular acts of copying involve two users or a greater number of users, the availability of a file logically depends on the decision of its original distributor to make it available. So the "events involving" people who share a file in a file-sharing system are equally "logically related" (or unrelated) in this sense, regardless of what technology underlies the file-sharing system.

I declare under penalty of perjury under the laws of the State of California that the foregoing is true and correct and that this document was executed in San Francisco, California.

Dated: May 27, 2011

By: 
SETH SCHOEN