UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF ILLINOIS
EASTERN DIVISION

Dallas Buyers Club, LLC,

Plaintiff,

v.

JIMMY KAWA, ELLIS QUALLS,
BEN LAWSON and BRYAN NARBERT,

Defendants

Case No. 1:14-cv-3517

DECLARATION OF DELVAN NEVILLE

1. My name is Delvan Neville. I am over the age of 21 and competent to execute this Declaration.

2. I am the owner of Amaragh Associates, LLC, a digital forensics company specialized in BitTorrent investigation. I am an ACE (AccessData Certified Examiner) as well as the author of a BitTorrent monitoring suite, EUPSC2k.

3. I was contacted by Scott Kane on April 11th of 2015 regarding monitoring I have performed in the past of BitTorrent swarms including those involving Crystal Bay Corporation (CBC) and the film Dallas Buyer's Club, and what the results of this monitoring says about the likelihood that a Defendant in this case was interconnected with other Defendants by nature of traffic for the same infohash with a short range of “hit dates” reported by CBC[1].

4. In summary, the likelihood that all of the Defendants joined here engaged in communication with even one of the other Defendants each is astronomically small, roughly 1 in 300 million. Further, even if that had occurred, Plaintiff could not produce any evidence to support this claim as their BitTorrent monitoring provider did not indicate support for the associated optional protocol (Peer Exchange) in communications with EUPSC2k, nor would all of the Defendants necessarily support this optional protocol either.

5. From mid-September through October 3rd of 2013, I performed BitTorrent monitoring and analysis work for the Electronic Frontier Foundation (EFF) meant to characterize the interconnectivity of peers within a swarm[2]. For these “soaks”[3], I monitored 24 swarms associated with IPP International-backed lawsuits, Crystal Bay Corporation (CBC) backed lawsuits, and swarms legally redistributing open-source software.

6. Though I had substantial pre-existing logs from soaks relating to both companies, I added new

[1] Although in this case the hit dates range over a period of just shy of 5 days, the statistics lead to the same conclusion even when all of the hit dates occur upon the same day.

[2] Here, the term “swarm” was used in the form that is typically used: to mean any and all users who have ever exchanged files for a given infohash on BitTorrent, which does not mean these users are a distinct interconnected group. This difference in meanings of “swarm” as used casually versus how it has been interpreted for purposes of joinder will be discussed in more depth.

[3] A “soak” is a continuous period of time during which one or more EUPSC2k nodes are connecting to peers exchanging the files associated with a given infohash, in order to monitor and record their activity.

features to EUPSC2k for the purposes of this work to allow deep analysis of Peer Exchange protocol messages.

7. Peer Exchange (PEX) is an extended BitTorrent protocol whereby, following a handshake message between two peers, the peers will notify each other of the IPs of all other peers they are currently connected to within the same swarm, and subsequently update in later messages when any of those peers have disconnected. The purpose of PEX is to allow swarm members to discover each other in addition to the use of one or more trackers and/or Distributed Hash Table (DHT).

8. Through the use of PEX, I was able to not only characterize how long a typical swarm participant remained as a leecher[4] and as a seeder[5], but with what percentage of the observed swarm any PEX-enabled peer contacted during their lifetime in a swarm. Although the inter-connectivity of peers who do not support PEX cannot be directly observed, it stands to reason that peers that do not support this optional method to find more peers in a swarm will at the most be as equally interconnected as PEX-supporting peers, if not less so due to non-PEX peers having fewer options for finding new peers.

9. During the first soak, which consisted of a day long monitoring of 17 swarms of either IPP-monitored, CBC-monitored or legal (thus presumably unmonitored) swarms, the average time spent in a swarm as a leecher was 0.996 hours and the average time spent as a seeder was 3.117 hours, though both distributions had standard deviations approximately 3 times the value of the mean, indicating that both leeching time and seeding time are highly variable on an individual basis.

10. Based on a record of all IPs detected in each swarm by an EUPSC2k node and PEX communication by the subset of peers who report PEX data, the average peer contacts only 0.61% of the total number[6] of swarm participants over the course of their time in the swarm, with a standard deviation of 1.35%. This indicates that a typical peer contacts only a sliver of all swarm participants, and while this distribution is also highly variable, 95% of swarm participants would have contacted between just a single peer to a maximum of 3.247%.

11. A second soak was performed on 7 more swarms, this time over a two-week period. This was directly inspired by mass-Doe litigation wherein the “hit dates”[7] would often be days or weeks apart, rather than consisting of Does present in a swarm on the same day. The findings for time spent in the swarm were similar to those from the day-long soak: the average download time was 0.603 hours, and the average upload time 2.042 hours. As before, the standard deviations were large, in this case much larger (over 6 times the mean for both average download as well as upload times). Percent connectivity was an order of magnitude lower, however, at 0.05% on average with a standard deviation of 0.15%. This finding was not surprising, if peers only remain in the swarm for an average total of less than 3 hours, it is extraordinarily unlikely that peers from one day will have communicated with peers on a second day, let alone peers separated in time by weeks.

[4] A “leecher” as used here is a member of a swarm who has not yet finished downloaded the contents of a torrent.

[5] A “seeder” as used here is a member of a swarm who has finished downloading the contents of a torrent, but is still connected with members of the swarm, typically in order to continue to share the file(s) with others.

[6] Here it is presumed that the swarm is limited only to those peers observed by one or more EUPSC2k nodes. The true number of peers exchanging that infohash could be much higher, and thus percent interconnectivity much lower, if a substantial number of peers were exchanging it on private trackers or obscure public trackers, or if by pure chance some trackers did not release information of all their known peers over the course of EUPSC2k accesses.

[7] “Hit date” is used here only to coincide with the terminology used in IPP/CBC exhibits, and is not meant to endorse the concept that a “hit date” is an appropriate way to describe how and when a peer participated in a swarm.

12. The results outlined above show that BitTorrent joinder litigation based is not based upon any real likelihood that the joined peers have engaged in any series of transactions with each other. The 4 peers whom are alleged to be Defendants in this case were separated by a period of two weeks (March 9th, 2014 to March 23rd, 2014), indicating the swarm would be interconnected as in the two-week analysis of paragraph 11 at 0.05% (the average peer in the swarm would have contacted 0.05% of the swarm during 2 weeks). The likelihood that there is any series of peer-to-peer connections that could link all 4 peers to at least one other named peer is 0.000000337%, or roughly a 1 in 300 million chance[8].

13. To discuss this organization in better depth, it is necessary to re-interpret the casual usage of "swarm" in technical discussion relating to BitTorrent to the actual organization of BitTorrent traffic. Swarm is casually used to be equivalent to "sharing the same infohash", but has been interpreted by some courts as meaning "a group of interconnected BitTorrent peers" and thus being part of the same series of transactions

14. A peer wishing to download (or upload) the file(s) corresponding to a given infohash seeks out other users through one or more of:

1) Tracker servers, which keep a short-term internal list of IPs that have contacted it seeking to exchange the files corresponding to an infohash, and responds to requests from a given peer with a handful of IPs & ports for a requested infohash.
2) The Distributed Hash Table (DHT), which accomplishes the same functions as a tracker, but where the list is stored in a distributed fashion in a cloud made up of any participant in the DHT rather than on a specific server.
3) Peer Exchange, explained above, which is a peer-to-peer "I'm actively connected to these peers right now for this infohash"
4) Ancillary methods such as manual entry by a user, or utilities for searching for a given file on another distribution network e.g. eDonkey2k

15. A BitTorrent file, identified by an infohash, gives a user an info key describing the files in the torrent, and an announce key describing some trackers the peer might contact. The "infohash" is generated *only from the info key*, and as such, two .torrent files with the exact same infohash can contain wildly different announce keys, and thus point users to entirely different trackers.

16. Any user may add to/remove from the list of trackers they choose to use for a given torrent, or even refuse to use any tracker at all by relying entirely on the DHT. Further, some trackers are private and require a password for access: users of these trackers may join distinct swarms (i.e. distinct groups of connected peers) exchanging the same infohash as on public trackers.

17. This distinction is one cause of the low interconnectivity among users exchanging the same

[8] This probability was calculated on the basis that any arrangement of communication that links each peer in this suit to at least one other peer would be sufficient. Peer #1 would have a 99.95% chance of not being connected to Peer #2 (100% - 0.05%), and so on for each other Peer. The chance Peer #1 saw none of the others is then 99.85% (99.95% raised to the 3rd power). Therefore, the chance that Peer #1 saw one or more of the others is 0.15% (100% - 99.85%). The likelihood that all Peer each saw one or more of the others is then 0.000000337% (0.15% raised to the 3rd power). This calculation omits the finite population correction factor, which would only become relevant when the number of Peer approached the total number of members in the swarm over that time period. Note that as the number of named Peer increases, the likelihood every Peer is linked to at least one other Peer *decreases* even though the likelihood that just one of all the named Peer is linked to a single other Peer *increases* i.e. a quasi-reversal of the Birthday Problem. In the language of the Birthday Problem, here we are not interested in whether there is just one pair of Peer with the same "birthday", but whether *all* Peer have a birthday in common with at least one other Peer.

infohash: EUPSC2k queries for new peers through all publicly available means, but most users query from a small subset of sources as only a few (or even a single) other peer is necessary for them to finish their download. Merely sharing the same infohash does not demonstrate that two peers were part of the same contiguous swarm.

18. As trackers only give out a handful of IPs & ports after each query, can hand out the same IP & port to the same peer on multiple queries, and only a very small sample of peers is necessary to finish most BitTorrent downloads, even showing that two peers utilized the same tracker does not necessarily demonstrate the peers were part of the same digitally connected group of peers exchanging a given infohash. This is especially exacerbated by the fact that tracker entries are extremely easy to falsify[9].

19. As the infokey is dependent on the contents of the file and the file name, the “original seeder” argument for joining Defendants on infohash is also troubled: any file that was first released through some means other than BitTorrent (including files legally acquired digitally from the copyright holder) can have multiple individuals indepedently choose to share it via BitTorrent. So long as they do no change the file name or the file's contents and leave the piece size settings at their default, they will generate the same infohash, and result in BitTorrent traffic that cannot be traced back to a single “original seeder”.

20. It should be noted that the IPs used by EUPSC2k nodes as well as the IPs used by IPP/CBC[10] were excluded from calculation of peer size as well as percent of swarm contacted by a given peer to ensure these passive observers did not artificially skew estimates of average connection times & connectivity.

21. As every communication between an EUPSC2k node and IPP/CBC demonstrate that they do not support PEX messages, even if the named peers in this case did engage in the same series of transactions together, Plaintiff will not be able to demonstrate that this occurred.

FURTHER DECLARANT SAYETH NAUGHT

Pursuant to 28 USC Sec. 1746, I hereby declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct and based upon my personal knowledge.

Executed April 13th, 2015

Delvan Neville
Amaragh Associates, LLC
570 NW Walnut Blvd
Corvallis, OR 97330-3849

[9] Because some users may use a proxy for outgoing traffic in an attempt to protect their anonymity, their outgoing IP seen by a tracker may be that of the proxy rather than the IP they use for incoming traffic. To solve this issue, trackers allow a peer to specify what IP and port to list, rather than only listing IPs that have *actually* contacted that tracker. This is easily abused to list any IP on the tracker without the knowledge or consent of any valid user of that IP.

[10] CBC's monitoring appears to actually be performed by IPP: the same IPs sending messages following the same exact script present in both IPP & CBC swarms in force, but absent from swarms unrelated to either company or their clients.