

Mary Jeka
Vice President

July 28, 2008

Honorable Nancy Gertner
United States District Court
District of Massachusetts
John Joseph Moakley U.S. Courthouse
1 Courthouse Way
Suite 2300
Boston, MA 02210

Re: Zomba Recording LLC et.al. v. Does 1-11
Case No. 1:08-cv-10895-NG
Subpoena dated July 9, 2008

Dear Judge Gertner,

As you are aware, Tufts University has received a subpoena from the Plaintiffs in the case referenced above. Although Tufts is not unfamiliar with RIAA's efforts to obtain the identity of alleged copyright infringers, the subpoena in this particular case contains distinct requirements that have given rise to some concern on the part of the University. We respectfully request the opportunity to discuss these concerns with you prior to disclosing the information requested in this particular subpoena.

In the lead case in this matter, London-Sire Records, Inc. v. Does 1-4, Civil Action No. 04-cv-12434, the Court's Order of March 31, 2008, adopted a new discovery procedure to be applied when an Internet Service Provider receives a subpoena requesting identifying information of a particular IP address but is unable to determine, to a reasonable degree of technical certainty, the identity of the user. In such an event, the ISP is required to submit to you, for an in camera review, a sealed list of all possible individuals meeting the plaintiff's criteria, as well as a brief statement explaining the reason it cannot determine which individual is the defendant. It is the submission of these names, and the ramifications of Tufts taking such action, that we wish to discuss.

As explained in more detail in the attached memorandum prepared by our office of Information Technology, Tufts uses two different processes to match an Internet Protocol (IP) address to a Media Access Control (MAC) address. One of the systems, Dynamic Host Configuration Protocol (DHCP), records the assignment of a specific address to a specific machine; this system is reasonably precise but records are maintained only for approximately ten days. The other system that we use when the ten day period has run under the DHCP system, Address Resolution Protocol (ARP), records entries from the University's routers at various intervals, but it only records the first time and last time that

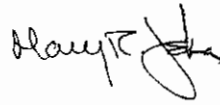
a particular user is assigned to a specified IP address within a given period. Using the ARP system may yield any number of users who had access to a specific IP address at some time before and after the time at which an alleged infringement took place, but it cannot tie a specific user to a specific MAC address at a specific time.

In the case of this particular subpoena, three of the specified IP addresses can each be matched to one user with a reasonable degree of probability (albeit less than total certainty), because only one IP address appears in our ARP data records. On the other hand, two other specified IP addresses are wireless access points which generally turn over more frequently; in one case there are twenty-three potential users who could have accessed the system at the time in question, in the other there are seventeen potential users. It is therefore difficult to conclude with any reasonable level of certainty that any one of those users was actually using the IP address in question at the relevant time. We believe, in these two instances, that it would be unfair to identify all possible individuals meeting the plaintiffs' criteria, given the low likelihood of identifying the guilty party.

We recognize the inherent limitations of the network data retention system that we are currently using, and are actively looking at possible adjustments. In the interim, however, we would like the opportunity to share with you the data that we have been able to gather, and allow you to make a determination as to whether the names of all forty individuals who could have had access to the two wireless points described in the preceding paragraph should be made available for further discovery. It is Tufts' belief, in order to comply with our obligations under the Family Educational Rights and Privacy Act (FERPA), that we are obligated to notify all students whose names are disclosed to you, even under an "in camera" inspection. As an alternative, producing the technical data for your review without providing the students' information (i.e., assigning each student a unique identifier) would eliminate the need to contact the students now, and spare them and their families from any undue distress in the event that you decide not to let discovery proceed.

We are available at your convenience to discuss this matter further.

Sincerely,

A handwritten signature in black ink, appearing to read "Mary Jeka", with a stylized flourish at the end.

Mary Jeka
Vice President
University Relations

Network User Identification Systems At Tufts University

Tufts University utilizes proprietary, legacy systems, built in-house for the management of the university's network. They were not designed to facilitate forensic investigations of user activity, but to ensure smooth operations and to manage capacity issues. Accordingly, there are limitations in the use of these systems to identify network users alleged to have engaged in copyright infringement. This document briefly explains (i) how our system works, (ii) what information is retained by the system and the points at which it is available for analysis, (iii) the limitations on those processes, and (iv) the resulting levels of certainty with which we can match Internet Protocol addresses to Media Access Control addresses to identify an individual user's actions.

Identification. Tufts University's system is entirely dependant upon a computer's Media Access Control (MAC) address and therefore identifies a machine, not an individual user. A MAC address is a hardware identifier for a computer, or, more specifically, for a computer's network interface card assigned by the vendor during manufacture. A laptop computer may have multiple MAC addresses, typically when it has both wired and wireless capabilities. When a computer is first connected to the Tufts network, in both wired and wireless instances, the user must "register" their computer's MAC address(es) with their individual username and password. During the registration process, our system records the MAC address(es) of the machine presented along with the username, and grants the user access to Tufts' system for one year. In many instances, a MAC address would be limited to one user. However, a public access terminal would have one MAC address, but multiple users.

Tufts uses the widely deployed – and industry standard – Dynamic Host Configuration Protocol (DHCP) to automatically assign or lease an Internet Protocol (IP) address to a computer's MAC address each time a computer on the Tufts network attempts to access the Internet. An IP address is a unique address that computers and devices use in order to communicate with each other on a computer network. Typically, an Internet Service Provider maintains a pool of IP addresses and assigns them to individual users as they access the network. DHCP electronically grants an IP address to a specific MAC address for a short and finite period of time and records the assignment. For as long as a user is maintaining its connection to the internet, its computer will attempt to continuously renew its IP address assignment; if the user logs off, the IP address is returned to the pool and the system may assign the IP address to another user. DHCP leases an IP address to only one computer at a time so they do not overlap. Information on when a "lease" was requested, its term or duration (typically six hours), and the MAC and IP addresses, are recorded in our system. So long as the user is continuously assigned to a given IP address, we can, with *reasonable* technical certainty, determine which MAC address was using which IP address at a given point in time. This is the preferred means of identification, but, as discussed in further detail below, there are technical and organizational limitations to this method.

In addition to the processes described above, Tufts uses a third system which records the association of a MAC address with an IP address via the Address Resolution Protocol (ARP). ARP is the method one device uses to find another in order for the two to communicate over a network. The ARP cache, or database, is a storehouse of tables maintained in networking gear, showing IP and MAC address matches with dates and timestamps. However, unlike DHCP, the ARP system has an inherent ambiguity. Because thousands of ARP records are produced each day, the system does not maintain a continuous or precise record, but rather records snapshots of data, showing the points in time when associations between given IP and MAC addresses were “first seen” and “last seen”.. The system does not maintain data on what may have occurred between the snapshots.

Retention. A user’s registration information (including its MAC address) is normally kept for several years. The DHCP data, on the other hand, is generally maintained for a period of ten calendar days from the date of issue by the system, after which it is overwritten by new data. The aggregated router ARP table entries are stored for an arbitrary period of time, varying from six months to many years, depending on the frequency of database updates. ARP data is the only long-term record that has the *potential* to connect a MAC address to an IP address but, as discussed above, it is imprecise and, even under optimum conditions, yields only circumstantial evidence of which MAC address was associated with a specific IP address at a given time.

Utilization of Data to Identify a Specific MAC Address. The ideal scenario for all parties is where a notice to preserve information comes to Tufts at a time that permits us, under normal working conditions, to preserve DHCP data while it is still available. In such a case, we are able to identify with reasonable technical certainty which MAC address was assigned an IP address that is alleged to have infringed upon a copyright. This was the case for three of the eleven alleged infringers named in the subpoena served upon Tufts University on July 9, 2008.

If an IP address specified in a preservation notice or subpoena cannot be matched to a MAC address prior to the DHCP information expiring, due to system or organizational resource constraints, any possible identification is limited to the imprecise ARP data.

Occasionally, only one MAC match comes up in the ARP database. Where a sufficient number of IP addresses are available to service the number of users in the system at any time (which is generally true for wired ports, such as dormitory rooms and offices), each time a user logs in, the network system will try to assign the user to the same IP address it used at its last log-in, assuming that address is still available. Therefore, if the IP address in question does not service a high volume of users, there is a reasonable probability that the single matched MAC address was, in fact, the computer in use at the time of the alleged infraction. This was the case for Does 4, 6, 9 and 10 identified in the subpoena. (Doe 4 has recently settled her case with the plaintiffs.) However, any such identification lacks the reasonable technical certainty of DHCP discussed above, since it is technically possible that another unidentified user accessed the system and used the IP address in question without being recorded in the ARP database.

If the ARP database provides multiple MAC matches to an infringing IP address, then a number of individuals (whose first recorded use of an IP address predates the alleged infraction, and whose last use postdates it) could be identified as potential culprits. However, the likelihood of identifying one of those users as the guilty party with any degree of certainty is extremely low. While we can identify which user's first and last use is most proximate to the alleged infringement, we cannot determine whether that user was in fact utilizing the IP address at the exact time in question; any of the other identified users could have acquired that IP address in the interim. Each of the IP addresses identified for Does 1 and 2 is a wireless access point with high turnover; there are 23 potential users (whose "first use" and "last use" bracket the alleged infringement) for Doe 1 and 17 potential users for Doe 2. Historically, for the reasons discussed herein, we have declined to provide such "potential" names based on our strong belief that the evidence available to us was inconclusive.

Finally, a third scenario sometimes arises, in which we perform a search using the information presented in a subpoena and none of the data available from any of our system's resources matches an allegation. In such instances, we are simply unable to produce any substantive information in responding to a request to identify an alleged infringer. This was the case for Does 3 and 8. In both cases the retention notice arrived in such close proximity to the expiration of the 10 day retention period for DHCP data amount of time that we were unable to access the data before it had been overwritten. We have not been able to find any technical reason why no information exists in the ARP logs but it is entirely possible that the date and time specified in the subpoena are inaccurate (this has occurred on at least one occasion in the past).

We are currently investigating the feasibility of altering our system to capture forensic data automatically upon receipt of a preservation notice, which may reduce our reliance on ARP data to identify an alleged infringer. However, we do not expect that such a modification would improve the operation of the University's system, and it would be implemented only at a substantial cost.

Appendix

Tufts ARP record limitations: A more detailed explanation

Each unique MAC and IP address combination is given one entry in the aggregated ARP data that includes only the date and time the unique pair was “first” and “last” seen on the network. Consider a simplified series of events: device A uses IP address 1 beginning on 4/2/2008 for 10 days, device B then uses IP address 1 on 5/3/2008 for 7 days. Device A then comes back online on 6/4/2008 and IP address 1 is available so it uses it through 6/8/2008. Finally, device B returns to the network on 7/5/2008 and uses IP address 1 through 7/12/2008. These events would be recorded in the aggregated ARP data as follows:

IP Address	MAC Address	First Seen	Last Seen
1	A	4/2/2008	7/12/2008
1	B	5/3/2008	6/8/2008

As a result of the consolidation of data it is impossible to tell who used the IP address between 5/3/2008 and 6/8/2008. Indeed, the only thing the data says with any accuracy is that device A used IP 1 on 4/2 and 7/12 and that device B used IP 1 on 5/3 and 6/8. Any other information must be inferred by the person interpreting the data and does not constitute a reasonable degree of technical certainty.

Examples of DHCP Data extracted from Tufts University Records

The DHCP records for a computer on the network throughout the day of June 12, 2008:

IP address	MAC address	DHCP server name	Lease start	Lease end
130.64.224.119	00:23:51:bc:f1:3e	dhcp2.medford.tufts.edu	2008-06-12 18:01:03	2008-06-13 00:01:03
130.64.224.119	00:23:51:bc:f1:3e	dhcp2.medford.tufts.edu	2008-06-12 15:16:50	2008-06-12 21:16:50
130.64.224.119	00:23:51:bc:f1:3e	dhcp2.medford.tufts.edu	2008-06-12 12:58:11	2008-06-12 18:58:11

Examples of ARP Data extracted from Tufts University Records

IP address	MAC address	First Seen	Last Seen
130.64.224.119	00:23:51:bc:f1:3e	2007-10-29 13:34:56	2008-06-12 18:58:11
130.64.111.218	00:13:02:d5:b3:50	2007-12-14 00:02:51	2008-01-15 23:12:45
	00:12:f0:48:fa:b6	2007-02-12 23:07:51	2007-12-20 19:22:46
130.64.111.218	00:13:02:bd:a7:d3	2007-12-19 18:52:51	2007-12-19 22:42:41
	00:16:6f:7c:78:49	2007-12-19 13:33:00	2007-12-19 18:42:53
130.64.111.218	00:14:a5:da:4d:62	2007-12-07 11:13:08	2007-12-19 13:22:58