



Monitoring and Identifying P2P Media

prepared for

SPEDIDAM

by

BigChampagne Online Media Measurement

09 January 2006



Introduction

The notion of a licensing system that would compensate rights holders for the sharing of their song recordings among online users has been discussed for several years by academics (law professors, economists), other interested organizations (Recording Industry Association of America, Electronic Frontier Foundation) and industry participants (various rights-holders, P2P software developers, etc.).

Much of this conversation has been informal, taking place at conference panels and in the press, but there are specific and ongoing studies and proposals that advocate different versions of this idea (see *References*). There is also significant opposition to the idea (at least in the United States) for many reasons.

The purpose of this document is not to examine or take a position on the legal or economic issues raised by the possibility of such a licensing system. We leave that to lawyers and economists. Rather, this document is meant to provide an overview of how online sharing of media can be monitored, identified, and those identifications matched to their rights holders for the purposes of payments to the rights holders *if such a system is implemented*.

As such, what follows is a presentation of the main technical considerations and components that would be factors in the implementation of the monitoring and identification aspects of a licensing system.

What Information Is Needed?

Before reviewing *where* and *how* media sharing can be monitored online, it is useful to define *what* needs to be reported.

As SPEDIDAM has identified, the following fields of information are required for each song:

- Title of the Work
- Name of the Main Performer
- Name of the Composer
- Record Label
- Year of the Recording
- Duration of the Work
- ISRC Code
- Number of Downloads

Included in this list is the information that would be required to administer a payment system to rights holders and also provide some basic statistical analyses of this activity for interested parties (activity by Performer, Composer, Record Label, etc.).



Of course, not all of this information is embedded in a given song file, especially in the context of P2P networks where files are generally created and edited by users and are not standardized to any format or set of required metadata. However, as long as the work can be identified using certain data points, the rest of the information can be compiled and provided in the reporting.

Geographic Location by Country

The country location of users (France) can be determined using the Internet Protocol (IP) address as a reference. The solution that BigChampagne employs provides coverage of over 99.99% of IP addresses, with accuracy rates “well over 99 percent” when matching an IP address to country.

Using this system, details of which are available in the *References* section below, the location of each user of a file-sharing system can be matched to a country.

No records of the IP address need be retained for reporting purposes. Rather, all that is required is a record of the country in which the IP address appears.

Identification of Works

There are three basic methodologies we can employ to identify songs shared online among users of P2P systems. Each approach has different characteristics, strengths and limitations, and it is important to note that these techniques can be implemented in combination. In other words, a system envisioned by SPEDIDAM or implemented by BigChampagne Media Measurement would likely use several of these approaches.

- **Audio Fingerprinting**

This category encompasses variations on a process that creates a mathematical representation of a particular recording, commonly known as an “audio fingerprint.” An audio fingerprint can be used to identify a specific recording despite the fact that different digital versions of the same recording may have very different characteristics. The fingerprints are small files which can be transmitted across Internet very quickly, even using low bandwidth i\Internet connections.

Audio fingerprinting can be very effective for identification, but there are two important factors in implementation. To generate a “match” an audio fingerprinting system needs to analyze mathematically each file and compare it to a library of known fingerprints for identification.

As discussed below, songs can be monitored and fingerprints generated at the user level or network level. Software that resides on a user’s computer (for example, software provided by a file-sharing network or by the Internet Service Provider, or ISP) could



generate fingerprints and send them to a centralized system for analysis and identification. Alternately, systems could be installed within the network (at the ISP level) to generate fingerprints and identify song files transmitted across the network in that way.

Either way, identification information can be sent to a central database for the purposes of tracking. This database would contain any additional information required about the work, like ISRC code and any other fields required, so that once the work has been identified, all key data fields identified by SPEDIDAM can be included in the reporting.

In some cases, ISRC codes may be embedded within the file itself in which case it would be possible to extract this information directly from the file. However, ISRC codes are not embedded in all digital song files and could not be solely relied upon for primary identification especially in the realm of P2P file-sharing, where close control over song formats and standards is ultimately impossible.

Note that all of this can be done anonymously, without any personal identification of the user.

- **Metadata Analysis**

In this approach, we use characteristics of a song file that can be determined without working directly with the file itself. With some variances by situation, examples of the type of information that can be observed across the network (without “touching” the file) includes file size, file type, file name, artist name, song title, duration, and hash (a mathematical representation of the specific file, though not the unique recording as with audio fingerprinting). ISRC codes might also be included (for example, MP3 format ID3 tags do support ISRC codes), but in this case we would rely on presumed accuracy of an editable “tag” and not a code embedded within the song file itself.

Some of the Information Available in MP3 Metadata Fields

Time	1/11/2005 0:16
File Name	10 Radiohead Motion Picture Soundtrack.mp3
Artist	Radiohead
Title	Motion Picture Soundtrack
Size	4,167,081
Hash	104C1032 86DE28....
Year	2000
Length	421
Quality	128

While this approach does not offer the same level of positive identification that audio fingerprinting can deliver, it also does not require that software be installed at key points to manipulate each file shared or downloaded. Once the Title of the Work and Name of



the Main Performer are identified, a database of corresponding information (such as ISRC code and other fields) can be used to fill in the required fields for reporting.

Metadata analysis supports a methodology of “passive observation” and, like the other methodologies, can be achieved without gathering any personal information about users.

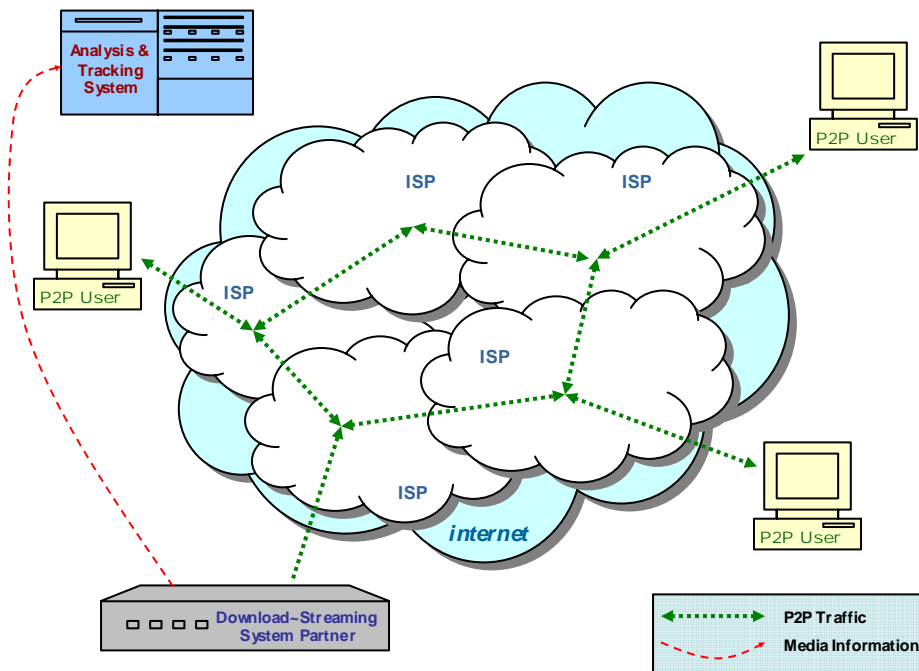
No records of the IP address need be retained for reporting purposes. Rather, all that is required is a record of the country in which the IP address appears.

- **Data Feed from Partners**

The third category of identification is based on cooperation with partners who already have a built-in ability to identify the songs on their systems. In this case, the partner organization provides an automated data feed which details the downloading activity on their network or system.

In this scenario, the network or system provides its own identification solution so the challenge then is to collect these data from multiple systems, analyze the raw feeds to generate download statistics, and package the information for reporting in a manner that is consistent with other sources and reports. Ultimately, a single, comprehensive report of activity is produced, regardless of source and identification methodology.

Fig. 1 - Data Feed from Partners



In this scenario, BigChampagne acts as a “systems integrator,” as discussed below.



Measurement of Activity

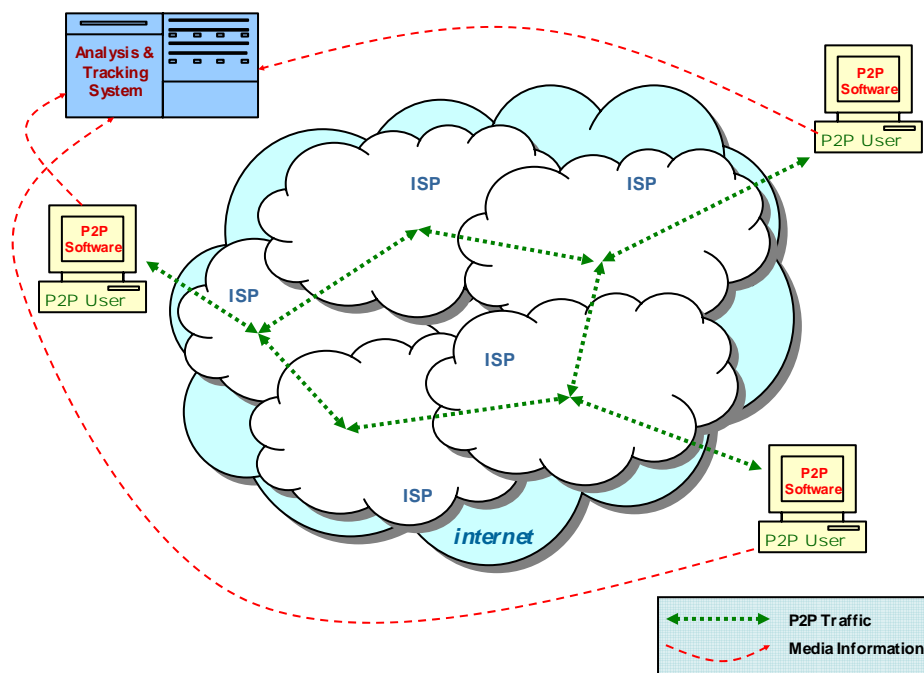
Bearing in mind the various methodologies for identification of song files, there are four basic approaches to monitoring network activity. Some of these have been referred to in the context of identification methodologies (above).

- **Measuring with the Cooperation of Software Developers**

This category is based on the cooperation of organizations and individuals that develop software used to access P2P file-sharing networks. Because these organizations control the software that users install on their computers to interact with other file sharers in a networked environment, they may be in a good position to observe what songs are being downloaded and to communicate that information to a central collection point.

Because the developer's software resides on the user's computer in this scenario, a modest amount of processing can be done at that local level without interfering with the user. Likewise, the user's Internet connection can be used occasionally to communicate basic information about downloads back to a central server without interrupting or bothering the user. Neither of these processes would affect the computer's performance, and they can be accomplished without using any personally identifiable information about the user.

Fig. 2 - Measuring with the Cooperation of Software Developers





A user's country location can be determined using IP address. No records of the IP address need be retained for reporting purposes. Rather, all that is required is a record of the country in which the IP address appears. Location might also be available from any registration or payment information that the developer may have for each user.

This approach is especially advantageous for the implementation of an audio fingerprinting system. The user's computer can generate audio fingerprints for song files as they are downloaded and their Internet connection can be used to send information about these audio fingerprints to a central location for analysis and reporting.

However, other types of identification can be implemented, also in cooperation with the software developers. They could, for example, simply transmit metadata about downloaded files to a central server or provide a data feed based on their own internal identification system.

The biggest challenge to this scenario is getting substantially *all* software developers to cooperate. Because the number of individuals who can create software for accessing file sharing networks is theoretically unlimited, it would be unrealistic to assume that *all* software developers would cooperate with the licensing system.

- **Measuring Activity in Partnership with Internet Service Providers (ISPs)**

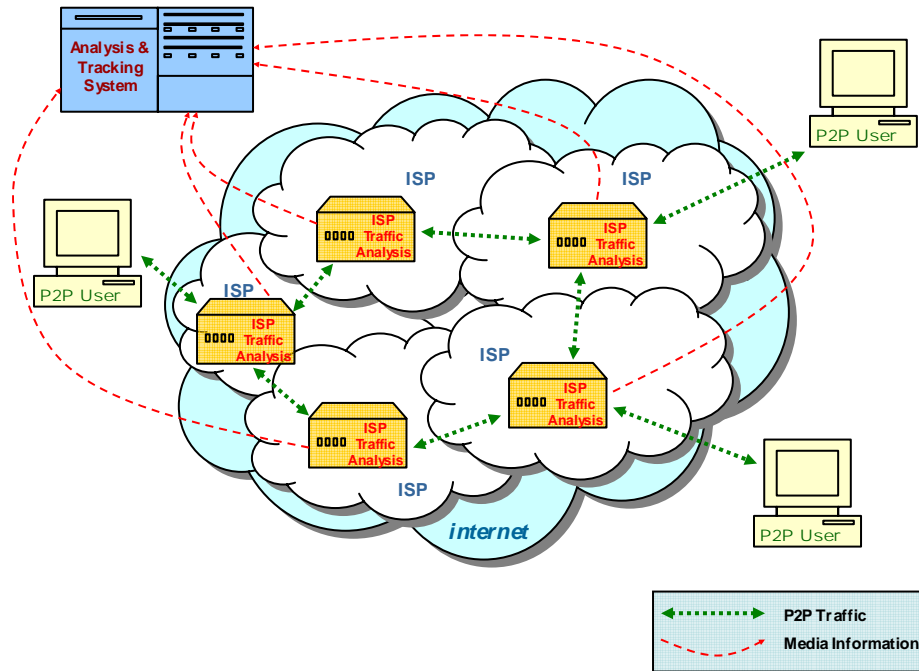
In this approach, measurement of activity is done in cooperation with ISPs rather than with P2P software developers as discussed above.

In this system, computers are connected to the ISP's network which can then examine all Internet traffic passing across the network. By directly monitoring Internet traffic, these systems can analyze the contents of the information being transmitted and identify individual songs.

User country location can be determined using IP address. No records of the IP address need be retained for reporting purposes. Rather, all that is required is a record of the country in which the IP address appears.



Fig. 3 - Measuring Activity in Partnership with Internet Service Providers (ISPs)



The resources required to implement this type of system would be significant due to the need for powerful computers installed at multiple points on each ISP's network. However, monitoring at the ISP level has the great advantage of bypassing any need for active cooperation by the software developers or users of P2P file-sharing systems.

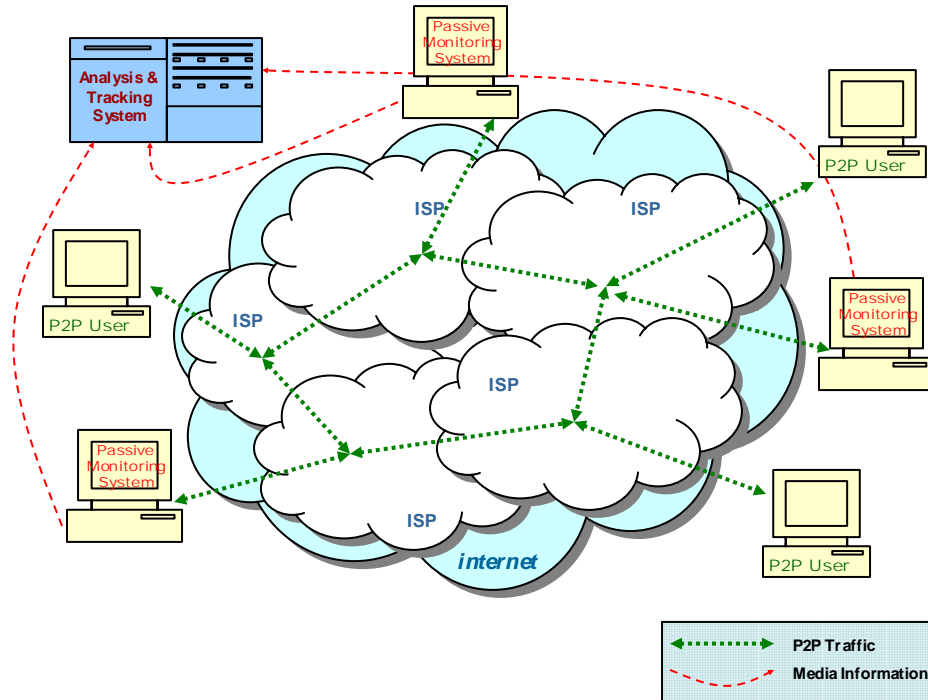
- **Passive Measurement without Partners or Permission**

The benefit of this approach is that P2P file-sharing activity can be monitored and analyzed without requiring any participation whatsoever. This is achievable because of the very nature of P2P file-sharing networks – they are designed to allow participants to search, browse and access the content shared by all other participants.

Therefore, it is possible for participants in the network to gather information about all other content on the network by participating in the network. By massively distributing this effort, passive monitoring systems can collect a vast amount of information about P2P file-sharing activity.



Fig. 4 - Passive Measurement without Partners or Permission



This methodology can employ a number of identification techniques including metadata analysis (metadata can be acquired easily in this scenario without acquiring actual files) and audio fingerprint analysis.

Due to the passive approach of this system, there are few privacy concerns. No software is installed anywhere, and no demands are placed on users' computing or network resources as in some other scenarios. User country location can be determined using IP address. No records of the IP address need be retained for reporting purposes. Rather, all that is required is a record of the country in which the IP address appears.

- **Measuring Using a Representative Panel of Users**

This approach is different from the other categories in that it does not seek to comprehensively measure all activity on P2P. In fact, this approach can be complementary to other methodologies.

Much like traditional market research, people can be recruited to participate voluntarily as part of a group whose song downloading activities are explicitly monitored. Users would be fully aware of this monitoring and could choose to stop being monitored at any time. This is similar to the way many U.S. households choose to participate as part of a



Nielsen panel for monitoring of television viewing habits (for the purposes of establishing television show ratings).

However, unlike traditional implementations of this type of “panel” research, this panel can be of immense size (there is no limit to how large this panel could be, and there is precedent for tens of thousands of households) and the monitoring activity would be automated. It would certainly make sense to implement an audio fingerprinting identification system in these circumstances, but any identification technique could be supported. Country location can be determined by IP address and/or through the process of registering a participant in the research group when they first choose to join.

Privacy concerns in this context are different because users would be giving explicit permission (“opt-in”) for the monitoring to occur and could choose to withdraw from the panel at any time. Even so, it still would not be necessary to retain any personally identifiable information about the users. While it would be useful for market analysis purposes to record things like gender and age range, even these are not necessary.

Role of the Systems Integrator

As should be clear from the above, given the great variety of file-sharing systems and download sources, there is no single approach to monitoring and identifying material that is optimal under all circumstances.

In some cases, cooperation with software developers and/or ISPs might be possible, while in other situations passive monitoring may be the only option. The possibilities are as varied and numerous as P2P file-sharing systems themselves.

As such, there is an element of the systems integrator that is required. A systems integrator should be able to monitor multiple sources, receive additional information from other systems or partners, and combine all of this together to generate comprehensive reporting of all activity.

The Systems Integrator must:

- **Standardize**
 - Accept multiple varied data feeds and standardize them to one format and system
- **Analyze**
 - Provide comprehensive statistical analysis of all of the activity
- **Audit**
 - Implement internal auditing mechanisms to ensure the integrity of source data and the resulting analyses
 - Accommodate third-party (outside) auditing of the data



- **Report**

- Provide ongoing and complete reporting of all activity as required by the agency administering payments based on these data
- As an option, provide tools to rights holders and other interested parties to view the information online. This could include not only basic information relating to payments, but could also include sophisticated market analysis tools for rights holders and others.

Ultimately, transparency would be critical. As a basis for distribution of payments to rights holders, all participants must have confidence that the system is open and fair.

In Conclusion

Critics of licensed music sharing rightly point out that any such system will be imperfect in implementation, and the many real challenges and limitations certainly should be considered. However, it should be noted also that the Internet affords measurement accuracy and efficiency previously unattainable and presently unparalleled.

Longstanding, established music licensing systems in the United States and around the world are often, quite simply, less scalable and less accurate than online media measurement. These systems, including those that compensate rights holders for traditional broadcasts (including radio, television, film, and even public performances) are now far from controversial. In fact, they have for decades provided the basis for remuneration of most creators, as they do today.

New technologies bring new challenges and threats, but also new opportunities. Proposed methods for measuring and monetizing media online are unrivaled in terms of both accuracy and scalability.

The revolutionary pace of the development of new broadband and wireless Internet capabilities virtually ensures that the distribution of popular music will continue largely undiminished. Parties on all sides of the resulting issues agree that artists, creators and rights holders must be compensated.



BigChampagne Online Media Measurement

BigChampagne Online Media Measurement is a privately held technology and market research company specializing in online media.

In October of 2003, *WIRED* anointed BigChampagne the Nielsen television ratings of online music. BigChampagne announced a strategic partnership with Nielsen Entertainment early in 2005, and has since announced data-sharing agreements with AOL Music, Yahoo! Music and MTV, among others.

BigChampagne pioneered the concept of peer-to-peer (P2P) measurement starting with the popular Napster community, and is today an industry-standard research tool.

BigChampagne's customers and subscribers include MTV/Viacom, major record labels, commercial radio stations, Wall Street analysts, artists, managers and other media industry professionals.

BigChampagne's chart syndication partners include Billboard Radio Monitor, Entertainment Weekly and E! Entertainment Television.



REFERENCES

A Sample of Works Discussing the Issue of P2P Licensing

- Neil Weinstock Netanel, “Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer File Sharing”
 - http://papers.ssrn.com/sol3/papers.cfm?abstract_id=267848
- William W. Fisher III, “Promises to Keep: Technology, Law, and the Future of Entertainment”
 - <http://cyber.law.harvard.edu/people/tfisher/PTKChapter6.pdf>
- Electronic Frontier Foundation, “A Better Way Forward: Voluntary Collective Licensing of Music File Sharing”
 - http://www.eff.org/share/collective_lic_wp.pdf

Geographic Location Identification using IP Address

- Digital Element’s NetAcuity system
 - http://www.digital-element.net/ip_intelligence/ip_intelligence.html