Before the

UNITED STATES COPYRIGHT OFFICE

LIBRARY OF CONGRESS

Washington, D.C.

_____

In the Matter of:	:	**Docket No. 2000-9**
	:	**CARP DTRA 1&2**
Digital Performance Right In	:	
Sound Recordings and	:	
Ephemeral Recordings	:	
_____	:	

TESTIMONY OF JONATHAN ZITTRAIN

INTRODUCTION AND PURPOSE

1.   My name is Jonathan Zittrain. I am the Jack N. and Lillian R. Berkman Assistant Professor of Entrepreneurial Legal Studies at Harvard Law School, where I study the intersection of law and technology, particularly Internet technology. I hold a JD from Harvard Law School, an MPA from the John F. Kennedy School of Government, and a BS from Yale University in Cognitive Science and Artificial Intelligence. I have been a "system operator" or SYSOP on the CompuServe Information Service since 1984. My current research includes understanding the means by which the Internet works, and the ways in which it is likely to evolve in the future.

2.   In this document, I seek to explain prevailing methods of real-time digital delivery of multimedia content, in particular the "streaming" delivery of digital music over the Internet, in which one can listen to the beginning of a piece of music even as subsequent portions continue to be received. I seek to document what I believe to be the essential characteristics of generally-used techniques. At the conclusion of this report, I describe modifications associated with alternative methods used to reduce costs and technical requirements, and I also describe cost and quality concerns affecting streaming today.

WEBCASTING - AN OVERVIEW

3.   The term "webcasting" is used to refer generally to methods of transmitting digital media (restricted in this discussion to digital audio) over a computer network, in general the Internet. Discussion in this document focuses on the webcasting of audio content, in general by radio broadcasters and by "internet-only" streaming services. Unlike person-to-person communications, such as videoconferencing or even simple telephone calls, webcasting generally follows a one-to-many architecture: the idea is for a particular piece of content to be made available to multiple people simultaneously, potentially even with each receiving the same experience at the same time. Unlike other means of distributing digital content, webcasting "client" software – the software installed on a consumer's computer – typically does not contemplate that the streamed content will be used by or available to the destination user after the stream has completed. In both these respects – the one-to-many design and the fleeting one-time-only playback – webcasts might be analogized to broadcast radio, as I discuss in detail at subsequent points in this document.

Purposes, Technical Challenges, and Notable Features

4.   Webcasting has developed to address certain problems previously faced by Internet users attempting to enjoy multimedia content retrieved from remote sites: First, the data representing a song or other multimedia experience can be large and thus time-consuming to transfer; webcasting attempts to address this concern by "streaming," i.e. by playing a portion of content while subsequent content continues to be retrieved, as well as by compressing content to reduce overall transfer requirements. Second, the pathways of the Internet can at times lose or delay data en route from source to destination; webcasting responds here with certain aspects of fault-tolerance, i.e. the ability to continue to play content even when portions of the data have not been successfully received – perhaps at a somewhat compromised level of quality from that intended by the webcaster.

5.   In both these respects, webcasting is substantially different from other widely-publicized methods of distributing digital media. For example, while content sent and received using software like Napster is generally compressed (most commonly in the

popular "MP3" format), Napster users generally must transfer ("download") an entire file before beginning to play it (lacking the "streaming" benefit of webcasts) and generally must successfully receive an entire sound clip in order to be able to play the file at the user's request (whereas webcasts can, in general, play a file even when portions are lost in transit). Of course, these benefits come at a cost: Sound quality is generally somewhat inferior in webcasts, as compared with content distributed via download in MP3 or other similar formats. I discuss sound quality in greater detail below.

6. The most widely used webcast systems currently available are the RealNetworks RealAudio and RealVideo package, Microsoft Windows Media Architecture, and Apple Quicktime. Discussion throughout this document applies to all of these packages, as well as to other similarly-designed programs.

User Experience Generally

7. A user's computer is in all lasting respects unchanged by the webcast process: In general, no new files are created; in particular, the process of listening to a webcast does not cause the capture of the content received to any disk file, and listening to a webcast generally does not add any new icons or other representations of content available for listening.

8. The webcast process described below is intended to be, and in general is, transparent to the user. That is, a user need not understand the steps described here; a user need only click and listen, without being knowledgeable about the special protocols and methods developed for webcasting.

STEPS IN THE WEBCAST PROCESS

9. The process of transmitting and receiving a webcast is characterized by a series of steps in a logical and temporal progression, as shown in the attached slide entitled "Overview." I here elaborate each of the steps shown in the figure.

Content Capture

10. As shown in the figure, the webcast process for music or other audio begins with the preparation for transmission of the content to be webcast (step 1 in "Overview"). The

source of the content to be webcast may already be digital (i.e. a CD) or may not (audio cassette, live content source). Either way, a representation of the sounds to be heard has to be transferred into a computer's storage mechanism – a process known as "ripping" when content originates on a CD or other digital medium and as "digitization" when content must be converted to digital form in the course of capture.[1]

11.   The result of this step of the process is that the content becomes a computer file, often in the uncompressed "WAV" format, generally stored in the permanent storage ("hard disk") of a general purpose computer in the custody of the webcaster, just as a traditional broadcaster might load a song or album into an automated jukebox – preparing to play that song as requested in the future.[2] Like all computer files, the content is ultimately represented inside the computer in binary form, meaning consisting solely of 1's and 0's, which must be interpreted by the computer in order to be meaningful to humans and their ears.

12.   At this stage, the original content (say, a song on a CD) has been copied onto a computer in what is generally a perfect or at least functionally interchangeable replica. But this perfect replica should not be confused with the encoded and compressed webcast content users ultimately receive from the webcasters; indeed, this perfect replica is typically not directly accessible by users of a webcaster's service, need not be retained after the further encoding in the next step, and in general is not retained since it is usually large; in fact, under certain methods of content capture and encoding,[3] the perfect replica in this stage is retained only in temporary working memory ("RAM").

---

[1] Discussion here and throughout is restricted to the webcast of prerecorded content rather than live sources. Webcasting live sources follows a somewhat modified approach at the Digitization and Encoding stages – the stages are combined, and performed on a single computer in real-time, followed by immediate transmission, again in real-time, of the streaming content to the streaming media server. In such instances, it is not necessary to store the digitized content on a computer's hard disk; the encoded content is retained only temporarily, i.e. in the encoding computer's short-term memory (RAM).

[2] Indeed, we note that in principle a webcaster could use a "live encoding" approach (as contemplated in the prior footnote) streaming content from a physical CD jukebox (i.e. a CD player holding multiple compact disks). While this method might be more expensive for the webcaster (have to buy an expensive CD jukebox), labor-intensive (staff to load and periodically update jukebox), and space-intensive (store large CD jukeboxes), it would achieve the same end result from the perspective of an end user of the webcaster's service.

[3] In particular, I refer to live streaming.

Encoding and Compression

13.   As part of the step 1 "preparation and transmission" process, content is next encoded in one of various webcast-ready formats (among them RealAudio, Windows Media, or Quicktime) in preparation for webcast distribution. Encoding is important to the webcast because it provides numerous cost and quality benefits, and is used for reasons similar to those of traditional broadcasters who modify their content to fit their medium, for example by limiting peak volume and by raising minimum volume. (This benefits listeners on the fringes of a traditional broadcaster's service area – for example, listeners who would otherwise be unable to hear soft passages due to the presence of static.) Similarly, the preprocessing benefits a webcaster's users who might otherwise receive an unacceptable transmission as a result of network unreliability. Again, note that these benefits come at a cost: a reduction in sound quality, as discussed below.

14.   Proper data encoding reduces transmission costs and requirements by sacrificing some sound quality. In particular, redundant sound data can be eliminated via data compression, essentially a kind of shorthand to reduce the amount of network capacity required to transmit content across a network. A full explanation of the methods of compression algorithms is beyond the scope of this document, but I seek to provide some sense of how it works (and can expand upon the subject in my oral testimony). Simplifying a bit, we might imagine within a song a prolonged period, perhaps a fraction of a second, of a single sound without variation. In such instances, compression might send the shorthand reference "continue this sound for 0.3 seconds" rather than the thousands of individual chunks of data otherwise required to represent that duration of sound. Actual compression routines are of course far more sophisticated, finding patterns in the many different sound elements simultaneously forming a single perceived sound, but the basic method is comparable. Encoding is also beneficial because certain widely-used methods of error-resistant encoding can improve quality in case of partial data loss during the course of transmission across the network.

15.   All commonly-used webcast implementations invoke a so-called "lossy" compression algorithm: accuracy is sacrificed to reduce network requirements. In particular, some sound content is permanently discarded in order to meet predefined

targets for transmission bandwidth ("bitrate"). While compression algorithms are designed to minimize perceived loss of sound quality due to compression, in most contexts at least some degradation is unavoidable, and even the ultimate decoding process on the user's computer cannot recreate the discarded data. Thus, the sound heard by the ultimate listener is in general somewhat inferior to the original content digitized at the content capture step. I discuss quality issues in greater detail in the section entitled "Resulting webcast quality," but for most webcast users, audio quality is comparable to that of an FM radio broadcast.

16.   For high bitrate encodings, some listeners say they can tell the difference between the uncompressed original and the encoded and compressed webcast version; others cannot. However, for lower bitrate encodings, including those webcasts suitable for users with standard dial-up modems, the reduced fidelity of webcasts generally is immediately noticeable. The quality of the user's audio components also has an effect here – after all, the sound emanating through a tinny computer speaker will likely sound bad whether or not the signal has been subject to lossy compression.

17.   In the future, it is possible that increases in available network capacity will render unnecessary this data compression, but in the short run such compression, and resulting quality degradation, is for most applications unavoidable.

Server

18.   In a typical webcast environment, each user receives her own independent transmission, despite the fact that others may be "tuned in" to exactly the same webcast and thus hearing the same thing. Such independent transmissions begin at the content provider's media server or servers – devices that generally store encoded content to be webcast[4] and, as users request content, initiate a separate transmission to each user. This process generally requires the use of specialized server software, and can require significant investments in server hardware.

---

[4] In a live streaming environment, the server instead receives a single transmission from the encoder. Note that the server may in fact be integrated with the encoder in certain implementations.

19.   This step of the process might be thought to be similar to the operation by a traditional broadcaster of a large broadcast tower, so long as one keeps in mind that, unlike webcast servers, broadcast towers are not dynamically burdened as increasing numbers of listeners tune in. Generally every additional listener to a webcast requires the webcaster to incur increased costs of signal delivery – costs not borne by broadcasters.

Transit and Intermediate Routes

20.   From the media server, content is sent on to those users requesting it (step 2 in "Overview"). As mentioned above, due to limitations in the methods of communication currently available on the Internet, in general a separate transmission must be sent from the server to each individual client. (These separate transmissions are shown in the slide entitled "Unicast.")

21.   This step of the process may be analogized to the transport of broadcast content from broadcast tower to listener's radio antenna through air waves. The one-to-many nature of the communication is similar to broadcasting insofar as the broadcast tower is capable of communicating with many radios simultaneously, just as a single server can communicate with many simultaneous listening users' computers.   Furthermore, just as a single transmission apparatus on a broadcast tower is capable of sending signals to many listeners' radios subject only to constraints of range, a single server can in principle serve arbitrarily many clients, subject only to the constraints of network design and server memory capacity. (Issues of bandwidth are discussed in greater detail in Variations on the Webcast Process, below.)

22.   In this stage of the process, the music content being transmitted over the network is not playable by any network operator or other intermediary en route for three different reasons.

23.   First, the content is not playable because it has been divided into segments too small to include a meaningful quantity of music content, and because each such segment has been removed from the context that would allow it to be played. (See the slides entitled "Data Extraction" and "Packetization.") Such division is necessary for transmission over the Internet; it is a fundamental principle of Internet architecture that

data transmissions are separated into tiny chunks ("packets") which are each separately transmitted from source to destination. The overwhelming majority of packets contain only raw data, not the contextual information necessary to produce even an instant of audio content. (By analogy, consider the difficulty of reading a piece of text if all vowels, spaces, and punctuation marks were removed.) As a result, even were it not for the compression process already described, and the interleaving process and multiple routing possibilities detailed below, the actual data being transmitted could not be repackaged into playable content at an arbitrary point in the process.

24.   Second, the so-called "packet-switched" design of the Internet makes possible, and indeed probable, that several packets of data sent from the same source to a particular destination will follow different paths to reach their intended destination. As a result, in general, no single device between the server and the user's ISP need receive all portions of the requested content; it might well be the case that half a song flows one way and half another, or that the various packets of data follow dozens or even hundreds of different routes in a period as brief as several minutes.[5]

25.   Third, the content is not playable at this intermediate stage of transmission because the encoding process has intentionally created "interleaved" files such that adjacent bits of transmitted data do not represent sounds from immediately-adjacent sections of music. (See the slides entitled "Data Extraction.") This approach is beneficial to webcast quality because it reduces the effects of loss or delay of any single packet; rather than a *particular* packet carrying the *entirety* of the audio signal for a particular snippet of sound, each packet carries *portions* of signals for *multiple* snippets of sound, allowing partial reconstruction of the sound even if some packets do not arrive at the player. However, since each packet thus carries bits of partial content from multiple snippets of sound, the data from any single packet would in general be insufficient to reconstruct any particular snippet of sound.

---

[5] Of course, in other cases, an extended series of packets might follow the same route for some time. In general, the closer (in terms of network topology) the content source is to its intended destination, the less variation takes place in packet routing.

Player/Decoder

26.   The requested packets of data next arrive at the user's computer and must be organized and processed somewhat before being played. Note again that data is "streamed" – that it arrives, is processed, and is played on an ongoing basis. Thus, an initial portion of a song arrives and is played for the user even before subsequent content is transmitted by the server or received by the user's computer.

Transport Buffer

27.   Since some packets may have traveled via faster routes than others, it is possible, in fact likely, that some misordering may have taken place en route; thus, the packets arrive in an order different from that in which they were sent. To address this problem, incoming packets are temporarily stored (for a few seconds) to assure that delayed packets are placed in proper sequence. Furthermore, one or more routes may have malfunctioned, causing some portions of data to become lost en route. In some implementations, the user's computer may respond by requesting retransmission of the lost data by the originating server, or if such retransmission is impossible or not timely, the user's media player may instead use mathematical algorithms to attempt to compensate for the missing information. In most implementations, the coordination of temporary packet storage and of requests for retransmission are managed by the operating system of the user's computer, in particular by its network protocols ("TCP/IP stack"), rather than by sound player software.

28.   As a series of packets completes this stage of the webcast process, the data stream has been stripped of any prefatory information indicating network transfer method, and the data has been reassembled into the encoded binary form in which it began its transmission across the network. Of course, the data still consists only of 1's and 0's, and it is still encoded using the data compression and interleaving previously described.

Media Engine

29.   The media engine is the portion of the sound player software that is responsible for converting the encoded data stream into a form capable of being played on the listener's computer. The data must first be decompressed according to the rules of the

chosen compression algorithm. Furthermore, any mathematical algorithms necessary to account for lost or missing data must be applied, to create a single continuous stream of digital information to be played back via the computer's standard playback routines. With this functionality, the user's computer is able to minimize the negative effects of lost or delayed packets of data on sound quality; however, if data was in fact lost (or badly delayed) en route, no mathematical algorithm can fully recover the missing information, and some reduction in quality is then inevitable. This process of decompression and decoding is shown in the slides entitled "Media Engine."

30.   At the conclusion of this stage, sound data returns to a form that can be played by standard routines. The data is then stored temporarily in short-term memory (for fractions of a second, and in small snippets) so that it may be converted into actual analog sound signals destined for the user's speakers as soon as preceding sounds are played, as is discussed below.

Audio Services

31.   At this point a snippet of sound data has been converted into a standard decoded uncompressed form. Thus, as shown in step 3 of the "Overview" slide, the standard audio services present in the computer's operating system can readily play the sound via the sound reproduction equipment connected to the computer. As additional packets of data continue to arrive, they are similarly processed via the transport buffer and media engine, and are subsequently in turn played by the audio services.

32.   As each snippet of sound data is played by the audio services of a user's computer, the memory previously storing that block is reallocated, becoming available for other use by the user's computer, most likely for use by new incoming sound data. In the implementations I am personally familiar with, at no point is the entire requested sound (or even a substantial portion of it) stored even in temporary computer memory (RAM); rather, as soon as each snippet of sound is played, the corresponding memory is reallocated to availability for the computer's other needs.

Sound Card and Speaker

33.   The output of audio services to the sound card is a continuous transmission of digital sound data in a standard, uncompressed, unencoded format. This data stream can be played by the sound card or sound chips installed in the user's computer. The sound card converts the digital signal back to analog form, yielding a standard electrical impulse capable of driving an ordinary speaker. One plugs headphones or a set of amplified speakers into the "sound jack" of the sound card of the computer, and the music plays.

34.   Sound quality can be further degraded at this step of the process for several reasons. First, computer sound chips can be of low quality, allowing electrical interference to taint the outbound audio signal. Second, low quality of cabling or of speakers may reduce fidelity of the ultimate reproduction of sound.

VARIATIONS ON THE WEBCAST PROCESS

Multicast

35.   When possible, it is desirable to send a single webcast signal to many users (as shown in the slide entitled "Multicast"). This dramatically reduces bandwidth requirements, lessening the costs of transmission. Furthermore, it noticeably reduces the burden on content servers, allowing the use of less expensive server equipment.

36.   Such methods are not widely used because multicast requires special additions to all network segments between server and user. In general, there is no guarantee that such additions are in place at all such steps, especially because routing can be unpredictable, as discussed above.

37.   For our purposes, multicast network design is notable because it may mean that all webcast data passes through certain key points in the network. However, data remains in encoded, compressed, and interleaved form at any such intermediate point.

Splitting

38.   When demand for a particular piece of content is such that no single server can satisfy all requests, it is necessary to configure multiple servers to share such requests. In general, one "primary" server sends signals to a predetermined set of other "secondary"

servers via a process known as "splitting" (as shown in the slide of that title). Splitting can be properly understood as the digital equivalent of multiple broadcast towers (perhaps in distinct geographic locations) transmitting the same signal.

Caching and Proxy Servers

39.   Whereas a content provider ordinarily would arrange splitting to reduce its own costs of distributing popular content, a network operator would have to take a somewhat different approach to reduce the costs of serving its users. In particular, consider an Internet Service Provider ("ISP," for example, AOL) which seeks to reduce the costs of providing externally-hosted content to its users. The content provider might choose to retain local copies ("cached copies") of the content at issue on a so-called "proxy server" on its own network. When its users request the popular content, the ISP satisfies the request using the file stored on its own proxy server, rather than re-retrieving the content from the remote content provider, thereby reducing the burden on the network operator's Internet connection to other networks.

40.   In general, the operation of proxy servers requires network operators to store and temporarily retain copies of webcast data. However, the data is generally stored on proxy servers in a way accessible to the network's users only via their standard player software and standard links. In particular, users listen to content hosted on proxy servers via the same process used to listen to content not hosted on proxy servers; the proxy servers are designed to be "transparent" to end users, and users are no more able to retain the webcast content when received via proxy server than they are when received directly from a streaming media server without an intermediate proxy.

41.   Splitting and proxy servers both address the same underlying concern by reducing bandwidth required to webcast content. For simplicity, most applications would use only one or the other, but the interaction of the two would not create any additional copies, temporary or permanent, beyond what is described above. Note that technical constraints prevent the use of proxy servers by those engaged in live streaming.

QUALITY, BANDWIDTH, AND COST CONCERNS

42.   To date, quality of webcast distributions has posed a significant challenge to content distributors and a disappointment to some users. I therefore consider these issues in additional detail here.

Determinants of Webcast Quality

43.   The quality of a webcast transmission depends primarily on two factors: First, quality depends on the "bitrate" at which content was encoded: Lower bitrates reflect the use of additional compression, but since compression is "lossy" (meaning that some content is discarded in order to achieve bitrate targets), applying additional compression reduces distribution quality. Second, quality depends on the success with which content is streamed across the network from server to client. If substantial portions of the content are lost (or delayed) en route, webcast quality will suffer, since the user's webcast player program will have to attempt to reconstruct the lost data.

Bandwidth and Connection Quality

44.   In principle, both quality problems can be solved via the use of additional bandwidth. If webcast content providers expect that additional bandwidth will be available from server to client, they can encode content at higher bitrates, allowing the distribution of a higher-quality original. Furthermore, if sufficient bandwidth in fact is available from server to client, the streaming of content from server to client will be more successful.

45.   Note that bandwidth alone may not be sufficient to solve some kinds of problems. For example, if data is being lost in transit from server to client (in the "transit and intermediate routes" described above), additional bandwidth will not improve quality. Rather, improvements would be necessary in the reliability of the overall transmission path, i.e. in upgrades to the intermediate network devices between server and client.

Cost and Availability of Bandwidth

46.   Webcast quality has proven a challenging issue to date in large part because increasing bandwidth is both difficult and expensive.

47.   From the perspective of content providers, webcast bandwidth is especially demanding because bandwidth requirements scale linearly as user-minutes of transmission increase. Since each user receives a separate transmission from the server, in a standard implementation, doubling the number of simultaneous users a content provider can support requires doubling the amount of necessary bandwidth. Furthermore, if each user requests webcasts of increased length, the total bandwidth required to support the users will increase proportionally. In the face of more users requesting ever-increasing durations of webcast content, content providers face a tradeoff between quality and capacity given a fixed amount of available bandwidth, and many have so far chosen capacity. In the long run, content providers can purchase additional bandwidth from telephone companies and Internet transit providers, but rates remain high, and availability may be limited or delayed in some areas.

48.   It is my understanding that a substantial portion of the costs of running a webcast are properly attributed to bandwidth. For example, the standard "T1" high-speed line used by many Internet system operators can support only forty simultaneous modem-connected users at generally-used bitrates, while such a connection to the Internet generally costs from one to two thousand dollars per month. Although costs decrease somewhat for larger implementations, it seems that webcasters' actual cost of distributing content remains substantial relative to other costs.

49.   From the perspective of an end user, obtaining an increase in bandwidth generally requires upgrading the user's on-site means of access to the Internet. In particular, because dial-up modems cannot receive high-bandwidth content, users of such devices must upgrade to alternative connection methods (in general, cable modems or DSL) in order to receive high-quality webcasts. However, such access can be costly, especially because it less readily shares other communication services users already have for other purposes (i.e. telephone lines). Furthermore, high-speed network access can require a lengthy installation process, and such access is not available in all areas or in all types of residences; it can be especially difficult to obtain high-speed access in rural areas and in non-owner-occupied housing. As a result of these difficulties, most individual computer users continue to rely on dial-up modems rather than high-speed alternatives.

Resulting Webcast Quality

50.   Given these constraints, webcast quality can be disappointing to some users. Dial-up modem users (as compared to those with high-speed connections) often find webcasts especially lacking: Audio distributed via webcast is generally radio-quality, while attempts to distribute video over modems yield small and jerky transmissions. Users of high-speed cable, DSL, or commercial Internet connections generally receive improved quality, potentially near-CD-quality audio, though data loss en route often continues to cause pauses, static, or other glitches in transmission.

Technical Issues and Potential Improvements

51.   While the alternatives presented in "variations on the webcast process" could improve quality dramatically while also reducing cost, they are as yet not widely implemented. Multicast is not widely deployed across the Internet, making its use impractical for most applications. Splitting can be difficult for service providers because a cost-saving implementation requires locating servers geographically and strategically across the Internet, generally in multiple countries, a task beyond many startup firms, although certain companies (including Akamai Technologies) specialize in assisting with this task. In the past, proxy servers have been considered an undesirable source of additional complexity in network design; thus, many users dislike them, and service providers often hesitate to install them. However, in the long run, it seems likely that progress will be made in implementing these three methods of improving quality while cutting costs; but, in the short run, improvements are likely to be difficult and relatively slow.

52.   Certain users are at times unable to receive any webcast content at all due to technical issues beyond their immediate control. In particular, users in certain shared installations (offices, libraries, schools, etc.) may face networks configured, intentionally or unintentionally, to prevent their retrieval of webcast content. Such blocking may result from the use of firewalls that fail to allow webcast data to pass through, or of client computers specifically configured, in order to conserve network resources or for some other reason, to lack (and to refuse installation of) webcast player software. When such

blocking is installed according to the best practices currently available, content providers are likely to be unable to transmit content to the users affected by the blocking.