

## [ I N S I D E ]

*Experts Say Napster, Playing Dumb, Resists Possible Tech  
Cure for Copyright Infringement*  
Charles C. Mann 10/18/2000 15:15

When news reports began filtering in about the Oct. 2 oral arguments in the Recording Industry Association of America's lawsuit against Napster, some members of the software programming community who know Napster's workings intimately were bemused. Members of the three-judge panel of the U.S. Court of Appeals, it seemed, were genuinely troubled by the prospect of holding Napster responsible for the alleged copyright infringement being committed by its millions of users.

Perhaps the judges were playing devil's advocates, but their questions about the company gave pause: "How in the world are they expected to have knowledge of what's coming off some kid's computer in Hackensack, N.J., for transmission to Guam?" asked Judge Robert Beezer at one point. Added Judge Mary Schroeder at another: "Napster doesn't have any idea at any point in time what's being transmitted and what isn't."

While it's hard to tell precisely what Beezer and Schroeder meant by the tough grilling they gave to RIAA lawyers, their apparent belief that Napster can't monitor or control its users echoes the company's public stance. "The underlying principle of the technology and the service is that... Napster provides the tools, but has no ability to impose limitations or exercise control," company co-founder Shawn Fanning testified before a Senate Judiciary hearing on Oct. 9.

But according to programmers who have reverse-engineered Napster's peer-to-peer sharing technology in order to develop open-source clones, the judges seemed to be discussing a different computer program than the one they are so familiar with. While Napster may not be able to control which MP3s the "kid in Hackensack" chooses to make available to the Napster system, programmers argue that the company could, with relative ease, screen out the great majority of infringing files from its music directory.

"Having been one of the many people who analyzed Napster's protocol, it's pretty clear to me that the judges were misinformed," says independent programmer David Weekly, a long-time fixture on the MP3 scene who has posted an analysis of Napster's protocol on the Internet. "Napster knows who is sharing what with whom else, and they could stop it."

Chad Boyda, co-owner and chief programmer of Thirty4 Interactive, a software outfit that produced Napigator, open-source software that allows people to search Napster and Napster variants simultaneously, agrees, saying, "Napster knows a great deal about what's going on in its servers. I don't know what they're talking about."

### 'IT'S THE LAWYERS' JOB TO TELL US THAT'

Napster's lawyers and experts, under questioning from the recording industry, have actually conceded that the company could theoretically exert some control. They have

protested, however, that such screening would inevitably be either too restrictive (filtering out songs whose distribution is authorized by their copyright holders) or too weak (filtering out only a tiny fraction of the available copies of any given offending song).

Of course, Napster also argues that it has no legal obligation to engage in such filtering, citing the U.S. Supreme Court's decision in *Universal City Studios v. Sony Corp.* In that case, the Court condoned the manufacture and sales of VCRs, notwithstanding the certainty that they would be used to infringe some copyrights, because they were also "capable of substantial noninfringing uses."

But setting aside the legal issue, Inside's own experiments and conversations with software experts suggest that use of such filtering techniques would very likely heavily pare down the company's libraries of unlicensed music -- so much so, that it would imperil its value as a commercial venture.

"It would really be very simple for Napster to do, to actually comply with the removal of identified infringing files," says Bruce Ward, technical director of NetPD, the startup in Cambridge, England, that monitors online copyright infringement and has intensively examined Napster, most notably for Metallica and Dr. Dre. He adds drily: "I don't know why they don't do it."

Steven Fabrizio, the in-house litigation chief for the RIAA, says: "What prevents Napster from stopping infringements is not technology, but their concerns for their business model. They are a business built almost exclusively on the most popular recordings in the world. If you take those away, 32 million people may not want to spend as much time on Napster."

In an interview, Napster CEO Hank Barry maintains that screening out allegedly infringing files is not yet technically "viable," though he said the company was constantly looking at the new possibilities that new technologies might create. In any event, he insists that Napster is doing everything that the Digital Millennium Copyright Act of 1998 requires it to do to honor the rights of copyright holders. "Respectfully," says Barry, "it's not NetPD's job to tell us what the law says. It's the lawyers' job to tell us that. We have to go with what our counsel tells us about whatever our responsibilities are under the law."

## THE MD5 CONTROVERSY

Napster's software has two parts, a "client-side" and a "server-side" in computer jargon. The client-side is the program that individual users download from the Napster's Web site; the server-side software resides on the company's computers at its headquarters in Redwood City, Calif. All requests for music go directly to the company's central servers, which send back the Internet address and full name of the requested file. In addition, the client on the machine with the requested file notifies Napster if it has accepted or rejected the request, sends periodic updates on the status of the uploading file and reports when the download is complete.

Importantly, the server also records the music file's "MD5 hash" -- a digital fingerprint, of sorts, that uniquely identifies the file. At its heart is an algorithm that generates a string of 32 letters and numbers that is unique to each MP3 version of a song. As a rule, any "rips" of a particular digital track made under the same conditions -- that is, the same software, the same settings -- should create files with the identical MD5 hash. (Click here for a fuller explanation.)

Initially, Napster used the MD5 hash for a "resume" feature. If a user logged off the service in the middle of uploading a song, Napster's servers automatically searched the central index for a file with the same MD5 hash. If one turned up, the servers would begin downloading it. Napster discontinued the feature last spring because, Barry maintains, it could not get the function to work properly. Programmers, however, speculate that Napster stopped using the popular feature due to legal concerns. The popular feature was discontinued at about the same time that RIAA lawyers had cited it in their briefs as evidence of why Napster was ineligible for the broad copyright immunity the DMCA provides for passive Internet Service Providers. (Open-source Napster clients like knapster and gnapster still offer that function.)

Record industry lawyers have argued all along that Napster could block unlicensed songs by a variety of means if they only wanted to -- by, for instance, simply screening file names for objecting artists and their songs before allowing them to be listed in the Napster directory. In court filings, Napster's outside expert J.D. Tygar, a Berkeley computer scientist, rebutted these suggestions. Tygar argued that filtering out certain file names would be overbroad, because many artists and songs have similar names, and some versions of songs, such as live performances, could be licensed for downloading even as other versions, such as studio recordings, were protected.

Filtering with MD5s, he continued, would be unworkable. Two users' "rips" of the same song could produce two files with different MD5 hashes, he said, because the files could vary in "the degree of compression, the exact start and stop points of the recording, and the device or software to create the MP3 file." And because many songs exist in several different versions, the number of disparate hashes would proliferate uncontrollably. "I have heard a number of CD recordings of Miles Davis's well-known album Kind of Blue that have dramatically different audio characteristics," he observed. "Similarly, I have heard multiple copies of Glenn Gould's 1956 recording of the Goldberg Variations with dramatically different audio characteristics." As a result, a single track by Davis or Gould could, in theory, generate an unworkably large number of MD5 hashes. In Tygar's view, sorting through the ensuing morass would be "technically infeasible."

## A DOWNLOADING EXPERIMENT

Tygar's objections may overstate the problem, however. Recently, Inside downloaded from Napster the first 30 copies listed in searches for 7 songs: "Old," the new Paul Simon single; "Shape of My Heart," the fast-rising song from the Backstreet Boys; "Optimistic," the lead radio track from the just-released Radiohead CD Kid A; "I Disappear," the Metallica cut whose appearance on Napster sparked the group's lawsuit against the

company; "Country Grammar," the ubiquitous Nelly hit; "Blue in Green," a classic track from Miles Davis's Kind of Blue; and Goldberg Variation No. 15 performed by Glenn Gould.

The results were surprising. Despite the theoretical possibility that thousands of different rips were floating around the Napster service, the overwhelming majority of what was available typically came from just a few rips. For instance, in 5 of the 7 cases, blocking just two MD5s would have screened at least 89 percent of the tracks. (For a table of the results, [click here](#).) Despite Tygar's prediction, there were just two MD5s of Gould's Goldberg Variation -- one of the 1956 release, and one of his considerably slower 1982 version. (Napster declined to have Tygar speak to Inside.)

Although the Inside test is only anecdotal, it's an accurate reflection of the overall picture, according to Ward of NetPD. Best known for its assistance to Metallica in the band's ongoing legal battle with Napster, NetPD has spent six months monitoring Napster and its clones, accumulating a database of about 40 million files that are available in file-sharing communities. (The company, long reclusive, received financing in August from UBS Capital, the venture-capital arm of the giant Swiss bank UBS AG; NetPD will debut a Web site this week.) From the monitoring data, Ward concludes that -- like Gnutella, in which a few generous souls apparently provide files for a mass of freeloaders -- Napster has a relatively small number of tech-savvy users who provide original rips and a vastly larger crowd of folks who copy those rips and pass them amongst themselves. The service, in Ward's view "is a kind of an echo chamber. Most of the time, it's a few originals and thousands of copies."

As an example, NetPD using proprietary software monitored three tracks -- "Optimistic" from Radiohead, Eminem's hit "The Real Slim Shady" and U2's new single, "Beautiful Day" -- for a 47-hour period between Oct. 7 and 9. The company found, respectively, 4,341, 16,090 and 8,077 copies of each song on Napster users' hard-drives. But the number of rips was far smaller. Just 10 MD5 hashes accounted for 94 percent of the copies of "Optimistic," 82 percent of "Slim Shady" and 95 percent of "Beautiful Day," the company reports.

Such figures convince Ward that Napster could effectively block out most infringing files, despite the theoretical problems raised by Tygar. If labels or band managers routinely sampled Napster for their material -- or, of course, hired a firm like NetPD to do it for them -- they could identify infringing MD5s that correlate to a recording and ask the company to refuse to allow files with those MD5s on its directory. When Napster users logged on, the company could quickly check their MP3s against a database of blocked tracks. If users had forbidden tracks, the database would simply keep them out of the company's directory, making them unavailable to other Napster users. In this way, Ward says, Napster could "greatly reduce" the level of unauthorized copying -- without having to kick a single person off the service.

(Both Metallica and Dr. Dre did, in fact, supply Napster with lists of MD5s for their songs that were being traded by its users, and demanded that the company block these MD5s from its directory. Napster refused, maintaining that doing so was not technically feasible and that it had no legal obligation under the DMCA to do so. Its only obligation,

the company said, was to disconnect the unlucky individual Napster users who happened to trade in Metallica or Dre songs during the period when those bands were watching and recording their actions. As a consequence, Metallica and Dr. Dre songs are still readily available on Napster.)

To make such screening possible, Napster would have to construct a database capable of rapidly searching huge numbers of MD5 hashes -- a task that it has called impossible. "Millions or even billions of (MD5s) must exist for all song files," Napster vice president of engineering Edward Kessler wrote in a July declaration. "Given the large universe of MD5 checksums, it is impossible for Napster to monitor the checksums when we process thousands of new files a second. Napster's service would be rendered unusable under such conditions."

But Steve Holzman, a representative of Unisys, a company that produces hardware and software for large-scale database, said "It sounds like an eminently feasible task" when the situation was described to him. Holzman, like other database-company representatives contacted by Inside, stressed that he was not addressing the specifics of Napster's system, but the general problem of creating a database that could constantly update and analyze millions or even billions of small entries. "People here don't regard that as terribly difficult," he says.

#### SEARCHING TERABYTES OF INFORMATION, NO PROBLEM

"Today's databases can search terabytes of data and provide sub-second response times," asserts John K. Thompson, vice president of worldwide marketing for WhiteCross Systems, whose Data Exploration Server is built for just such operations. (A terabyte -- a thousand gigabytes -- is roughly equivalent to 60 billion MD5 hashes.) "You could either perform the checking of the hashes on a central server that would provide millisecond response time or you could distribute smaller (slightly slower) systems in a number of locations in the network," he says. "You put one in the U.K., one in the U.S., one in Asia. As people come into the system, you constantly update the centralized or distributed databases." Spreading the load, he says, "reduces the latency associated with the network -- the people on the service won't even notice the database is there."

Apprised of such comments, Napster's Barry expressed interest. "I'm sure that's right in principle," he said. "How you turn that into something that works in a manner that's useful on a day-to-day basis is a real challenge."

"What we've got now is a good balance of a bunch of different concerns. We get suggestions all the time," Barry said with a laugh, "for how we might improve the system. We look at them."

Still, even such a database of MD5s would not be a cure-all for the record companies and artists that are suing Napster. Constantly monitoring the service for infringing rips would be a burden for labels and managers, much like the burden borne by retailers who must constantly be on the alert for shoplifters. And security experts caution that hackers will soon create versions of Napster software that scramble MD5 hashes -- "it'd

be easy to write software that would do it automatically," argues Dan Farmer, an Earthlink security researcher who is one of the labels' expert witnesses in the Napster case. Yet for such programs to have a wide enough impact to overwhelm a large, modern database, they would have to rise out of the underground and be embraced by millions of ordinary Americans.

Officials with five of the many would-be "legal" Napster clones now preparing to launch hold different views, sometimes strikingly so, about the potential value of MD5 screening. Flycode, for instance, plans to use MD5 filtering along with other proprietary technologies and partnering with an encryption firm, according to a spokesperson. So does Pointera, which according to founder and marketing VP Manish Vij intends to implement MD5 screening early next year.

But iNoize -- a streaming, as opposed to downloading, service -- does not plan at the moment to use MD5 screening. Nonetheless, the company's technical director, Alistair Fraser, says that it would not be difficult to screen out "99 percent of (infringing) file transfers by looking for song names and then checking individual MD5 hashes." Clarence Kwan, CEO of Lightshare, says his company, finding that MD5 screening would not be foolproof, has opted to screen files with its own proprietary technology, while also partnering with an encryption firm. Finally, Adam Powell of Angry Coffee -- which hopes to persuade the record industry to license its catalogs for free distribution in exchange for providing detailed information about consumers -- predicts that any form of screening will ultimately fail.

In any event, there is another, seemingly air-tight variant of screening which apparently avoids these problems altogether: Napster could limit (or be ordered to limit) its directories to tracks whose MD5s have been explicitly cleared for free distribution by their copyright owners. In that case, its peer-to-peer system of distribution would still exist, but could be limited to the songs of Napster's roster of unsigned "New Artists" and those of any other artists who wanted their work distributed that way. Whether such a business could ever make money is, of course, another question.

#### PATEL, IN RETROSPECT

Last June 26, when Federal District Judge Marilyn Hall Patel issued her order -- since stayed -- enjoining Napster from facilitating copyright infringements, she commented from the bench: "Plaintiffs have argued, and I think persuasively, that defendant is capable of exercising supervisory powers over its service. Though it may be technologically difficult, I'm sure that anyone as clever as the people who wrote the software in this case are clever enough (to) come up with a program that will help to identify infringing items as well. I think the evidence shows that there's no desire to do that."

Within hours of hearing Patel's comments, Napster attorney Jonathan Schiller took to the airwaves to assert that Judge Patel had simply failed to understand how Napster works. But based on the views expressed by the peer-to-peer software experts who have spoken to Inside, that does not appear to have been the case. Napster's problem may have been that she understood all too well.

With additional reporting from Warren Cohen.