

How to build a business around  
wholesale copyright infringement  
... without getting sued!  
*or,*  
minimizing your Usenet server's exposure to liability.

Joseph Barillari<sup>1</sup>

13 May 2003

<sup>1</sup>I pledge that this paper represents my own work in accordance with University regulations.

# Contents

## 0.1 Social background

From every corner and every pulpit, rabble-rousers in the employ of music-makers are crying out to the musicloving masses about a murder in progress. But like the bovine Queensites who rubbernecked as Kitty Genovese<sup>1</sup> was murdered, they remain unmoved by the shrieks and wails from Cary Sherman. and Jack Valenti.

Only this time, the lethal knife plunges not towards a hapless young woman, but towards a content-distribution model, and, perhaps, an entire industry. And this time, the bad Samaritans are not merely sitting in their apartments, chewing popcorn and watching for the blood, but joining in the murder in earnest. Each time a Morpheus or Gnutella client boots up, one more nail sinks into the business model's coffin. Each time a user hits the "Download" button in eDonkey, Vivendi Universal is dragged one step closer to the 'charnel house.'<sup>2</sup>

A cadre of curiously-named software programs have taken up the standard of their fallen comrade and founding father Napster. KaZaA, Grokster, Morpheus, eDonkey, Direct Connect, and Gnutella (to name a few) are the tools of choice of the broadband belt's light-fingered audiophiles.

The public watches, disinterested, as the life's-blood of CD-and-DVD-purchasing-income trickles from Jack and Cary's bodies, first through narrowband modems, then through the large-bore needles of college networks, and now, through innumerable fat pipes into cable and DSL subscribers homes, the trickle becoming a roar, and the industry hallucinating from

---

<sup>1</sup>[http://www.crimelibrary.com/serial\\_killers/predators/kitty\\_genovese/](http://www.crimelibrary.com/serial_killers/predators/kitty_genovese/)

<sup>2</sup>So runs the famous "BSD is dying" Slashdot post. <http://bsd.slashdot.org/comments.pl?sid=33311&cid=3598117>

income-loss, issuing statements that would make Mohammed al-Sahaf blush,<sup>3</sup> striking out wildly against their public, but ultimately failing to stem the flow. The public? They're too busy telling Madonna *precisely* what they think they're doing to care one whit for the music industry impending demise.<sup>4</sup>

New tactics emerge. The MPAA and RIAA<sup>5</sup> appeal their cases to the people, bombarding them with instant-messages<sup>6</sup> — a shoestring-budget sort of PSA, where the message is “DON'T STEAL MUSIC.” Apple's online music store puts the “we can't compete with free” crowd to shame by *selling* over a million songs in its first week.<sup>7</sup> (On a Mac-only service, mind you.) Grokster and Morpheus win an astonishing summary judgment for their right to exist.<sup>8</sup> And industry lobbyists carry on a steady agitation for laws mandating that every new computer be hobbled with copyright-enforcement technology — a permanent trigger-lock of sorts for the digital age.

Joe and Jane College probably didn't pay much heed when the jackals were set upon Dan Peng and his fellow search-engine operators, nor did they notice when they dropped him, unscathed, albeit \$15,000 lighter<sup>9</sup>. The siren song of free Christina and free Justin and free Britney is simply too seductive to dwell upon a few unfortunates who neglected to maintain minimum safe distance when the RIAA's lawyer-alarm went off. And why should they dwell? RIAA certainly doesn't plan to sue *them*.<sup>10</sup> The lawyering will go on, but so will the downloading.

---

<sup>3</sup>The best example of which came in 1982, well before the current crisis: <http://slashdot.org/articles/02/05/31/1622232.shtml> . A wonderful recent example is recorded at <http://www.arstechnica.com/archive/news/1040056422.html>: a RIAA press release indicated that a “piracy bust” netted “the equivalent of 421 CD-R burners.” It was later revealed that the number was less than half that. According to the RIAA: “There were only 156 actual burners, but some run at very high speeds: some as high as 40x. This is well above the average speed[.]”

<sup>4</sup><http://news.zdnet.co.uk/story/0,,t269-s2133749,00.html>

<sup>5</sup>Motion Picture Association of America (a cartel of moviemakers) and the Recording Industry Association of America (a cartel of musicmakers)

<sup>6</sup><http://www.mtv.com/news/articles/1471624/20030430/index.jhtml?headlines=true>

<sup>7</sup><http://www.apple.com/pr/library/2003/may/05musicstore.html>

<sup>8</sup>[http://eff.org/IP/P2P/MGM\\_v\\_Grokster/030425\\_order\\_on\\_motions.pdf](http://eff.org/IP/P2P/MGM_v_Grokster/030425_order_on_motions.pdf)

<sup>9</sup><http://www.dailyprincetonian.com/archives/2003/05/02/news/8154.shtml>

<sup>10</sup>“There is no next step. We are just letting them know it's illegal and they are not anonymous.” —Cary Sherman, RIAA president, <http://asia.reuters.com/newsArticle.jhtml?type=entertainmentNews&storyID=2655068>

## 0.2 How Usenet Works

Meanwhile, Usenet plods on as if it were 1990. Gaining nary a mention in the mainstream press, few casual net users are even aware it exists.

Usenet is a global discussion forum, the “largest decentralized information utility in existence.” Usenet is ancient. It predates generations of Internet users, predates the popularization of the World Wide Web, predates the Web itself.<sup>11</sup>

Mechanistically, Usenet is a collection of machines that have agreed to share “news” articles. The articles are organized into discussion groups focusing on a single category, termed “newsgroups.” Newsgroups are named hierarchically: for instance, the group `comp.os.linux.advocacy` is rooted in `comp`, the category for computer-related discussions. It is focused on computer operating systems (`comp.os`), particularly the Linux operating system (`comp.os.linux`), and it is further focused on Linux advocacy (`comp.os.linux.advocacy`).

In all but the tiny handful of moderated newsgroups (where an editor decides what will be posted and what will be suppressed), anyone can post an article for all of a group’s subscribers to read. Every day, thousands do, celebrating what Cliff Stoll lovingly called “anarchy in action.”<sup>12</sup> The lack of central control means that the network has seen its share of spammers,<sup>13</sup> trolls,<sup>14</sup> pathological flammers,<sup>15</sup> miscreants, and pranksters.<sup>16</sup> But many, if not most, posters are legitimate discussants who ask and answer questions, air ideas, offer criticism, and generally engage in civil conversation.

Of late, and due in no small part to the proliferation of high-bandwidth links to end-users and ultra-high-bandwidth links between news servers, Usenet has become a conduit for file-trading. Images, movies, music, and software are encoded, split into chunks (to avoid tripping post-size limits) and posted to special-purpose newsgroups, often called binary groups (“bi-

---

<sup>11</sup><http://www.catb.org/esr/jargon/html/entry/Usenet.html>

<sup>12</sup>Cliff Stoll, *The Cuckoo’s Egg*

<sup>13</sup>Those who post unsolicited advertisements, or “spam” in Internet jargon.

<sup>14</sup>Those who post “flamebait” to start a “flamewar” argument – for instance, posting one’s favorite beef recipes to `soc.religion.hindu`

<sup>15</sup>Those who can’t resist starting or joining into a good, long, flamewar.

<sup>16</sup>[http://www.wired.com/wired/archive/2.05/alt.tasteless\\_pr.html](http://www.wired.com/wired/archive/2.05/alt.tasteless_pr.html) describes an “invasion” of `rec.pets.cats`, a group for cat-owners by `alt.tasteless`, a group dedicated to the continuation of childhood gross-out contests. Thanks to Google, such historical flamewars are recorded for posterity: <http://groups.google.com/groups?threadm=132315Z24081993%40anon.penet.fi>

nary” data is any non-text data, like movies or music<sup>17</sup>). Usenet carries hundreds (possibly even thousands) of binary groups,<sup>18</sup> which range from the expected `alt.binaries.mp3` to the innocuous `alt.binaries.clip-art` to the praiseworthy `alt.binaries.missing-kids` to the astonishingly specific `alt.binaries.pictures.erotica.fetish.female.socks` and to the truly bizarre

`alt.binaries.pictures.erotica.anything.but.tuna`.<sup>19</sup>

If Apple’s Music Store and eMusic.com are clean, well-lit, legal emporia, and KaZaA and its ilk are the vaguely shady used-parts vendors at a computer convention, Usenet file-trading is the seedy black-market of a third-world nation. Usenet file-trading is the back-alley of the ’Net: steamy, dirty, filthy, and not just vaguely illicit — but downright illicit. It may indeed be the visual design of the KaZaA client that accounts for the public’s lack of moral qualms about file-trading:<sup>20</sup> the clean, airbrushed panes go a long way towards legitimizing the downloads; just as sifting through the disorderly messages on `alt.binaries.mp3` leaves one feeling in need of a hand-washing.

Articles posted to Usenet are ephemeral. Each Usenet site elects to keep them for a period — ranging from days to months to eternity. In low-volume discussions, this “retention period” is longer because the archives can cover a lot of time in very little disk space. In high-volume binary groups, most sites elect to cut retention time to a matter of days. Keeping the news any longer, even in this age of ever-dropping disk prices, is prohibitively expensive. For instance, Supernews.com, a commercial Usenet service, reports that its servers retain `alt.sysadmin.recovery` articles (which are typically text) for 304 days, and `alt.binaries.mp3` articles (which are typically binaries)

---

<sup>17</sup>Information theorists may object to this terminology, because text messages are also fundamentally represented in binary. This misuse of the term is, alas, deeply embedded in Usenet terminology.

<sup>18</sup>The “Authoritative active file,” a comprehensive list of Usenet newsgroups, lists 41187 groups, 2153 of which have the word “binaries” in their name.

<sup>19</sup>As an aside, the `alt.` hierarchy is the *most* anarchic part of this anarchic network: anyone can issue a command to create an `alt.` group (anyone can destroy one, too) — a command which site administrators worldwide are free to acknowledge or ignore. Contrast this to the “Big 8” hierarchies (`comp`, `humanities`, `misc`, `news`, `rec`, `sci`, `soc`, and `talk`),” where group creation follows a lengthy discussion and voting process. (<http://web.presby.edu/~nnqadmin/nnq/ncreate.html>)

<sup>20</sup>A plurality believe that there’s nothing wrong with music-sharing: <http://www.thestandard.com/article/display/0,1151,18987,00.html>

for 1.6 days.<sup>21</sup>

Some sites retain Usenet content forever. The best-known is Google Groups<sup>22</sup> (formerly DejaNews), whose searchable text-only archives stretch back to 1981.<sup>23</sup>

Guba.com (the “Gigantic Usenet Binaries Archive”) also archives Usenet content. But to a first approximation, it is the polar opposite of Google Groups: whereas Google archives no binary content, Guba archives *only* binary content. The aim of this paper is to analyze the legality of sites like Guba.com. To that end, I will explore the legal issues involved in running an ordinary Usenet server, then discuss the legal issues surrounding Google Groups. I will then use both to discuss the legality of binary archives like Guba.com.<sup>24</sup>

### 0.3 The legal status of vanilla Usenet servers

The chief legal issue facing Usenet is copyright.

It is not infrequently that copyrighted content is posted to Usenet without the consent of its copyright-holders. Individuals who make these posts expose themselves to charges of direct copyright infringement, as do those who download them. For brevity’s sake, this paper does not discuss individual users’ liability, but only that of Usenet server operators.

Authors who post to Usenet implicitly consent to the dissemination of their own words through Usenet. For this analysis, we need only consider cases where a user posts an article with content over which he has no fair use claim, against the wishes of its copyright-holder.

We will consider three varieties of copyright infringement (one primary and two secondary) for which a service provider may be liable in the course of this analysis. An Internet service provider (ISP) can be liable for **direct infringement** if they usurp the “exclusive rights of the copyright owner”<sup>25</sup> (by making copies of of a work without the owner’s consent, for instance). ISPs can also be indirectly liable for direct infringement committed by their

---

<sup>21</sup><http://www.supernews.com/stats/retention.cgi?group=alt.sysadmin.recovery&type=Exact> and <http://www.supernews.com/stats/retention.cgi?group=alt.binaries.mp3&type=Exact>

<sup>22</sup><http://groups.google.com>

<sup>23</sup>[http://groups.google.com/googlegroups/archive\\_announce\\_20.htm](http://groups.google.com/googlegroups/archive_announce_20.htm)

<sup>24</sup>I selected Guba.com as a concrete example, much of this analysis applies to Guba’s siblings, such as 2jacks.com.

<sup>25</sup><http://www4.law.cornell.edu/uscode/17/501.html>

users if said infringement is committed with their knowledge and assistance. This doctrine is called **contributory infringement**. A third variety of infringement, called **vicarious infringement** holds ISPs (or third parties in general) liable for any infringement that they had the power to stop, so long as it brought them financial benefit (even if they were unaware of the infringement).

The much-maligned Digital Millennium Copyright Act of 1998 (hereafter “DMCA”) shields ISPs from liability for carrying infringing content on their networks (in certain limited cases) so long as they follow a few rules which boil down to removing infringing material that comes to their knowledge and designating a representative to receive notices of infringement from content owners who discover copyrighted material floating on an ISPs network.

The four “safe harbor” provisions in the DMCA<sup>26</sup> shield ISPs from liability for material transmitted by users over their networks, material saved by automatic caching systems, material to which automated search engines link, and material uploaded by users. With some explanation, properly designed Usenet services are protected by the first of these.

Usenet servers obviously aren’t search-engines, nor are they filesystems solely under users’ control. At first glance, they sound like caches (because they take articles from the outside world, and hold them for their users to browse and download). But on closer reading, the system caching exception does not protect Usenet servers.

### 0.3.1 The “system caching” safe-harbor does not apply

(b) System Caching. -<sup>27</sup>

(1) Limitation on liability. -

A service provider shall not be liable for monetary relief, or, except as provided in subsection (j), for injunctive or other equitable relief, for infringement of copyright by reason of the intermediate and temporary storage of material on a system or network controlled or operated by or for the service provider in a case in which -

A typical Usenet server configuration stores articles temporarily, functioning as an intermediate party between the poster and the reader, as described in the information on transitory communications below.

---

<sup>26</sup><http://www.chillingeffects.org/dmca512/>

<sup>27</sup><http://www4.law.cornell.edu/uscode/17/512.html> #512.b

(A) the material is made available online by a person other than the service provider;

Subparagraph (A) requires that the cached material be posted online by someone other than the ISP. Nearly all news articles (barring administrative announcements, which generally don't spark copyright lawsuits) are posted by people other than the ISP, so this provision is easily met.

(B) the material is transmitted from the person described in subparagraph (A) through the system or network to a person other than the person described in subparagraph (A) at the direction of that other person; and

Subparagraph (B) stipulates that the material be transmitted through the ISP by the poster's system, at the request of the reader's system. This isn't the case with news: in the general case, the ISP's news server software automatically pulls articles from its newsfeed and makes them available to the ISP's users. The ISP-side user (the "other person") only comes in later, when she pulls the news files from the ISP's news server.

Unless the "other person" can be an agent of the ISP, Usenet doesn't meet this stipulation. It seems quite unlikely that the requesting user could be part of the ISP, for if that were true, it would collapse the ISP/user distinction articulated throughout the safe-harbor provisions. Nevertheless, it is unusual that subparagraph (A) explains that the *poster* must be distinct from the ISP, but there is no similar provision regarding the requester. We will assume, for the purposes of this argument, that the *requester* must be distinct from the ISP. Consequently, the caching safe-harbor does not apply. (There are more conditions on the system-caching safe harbor, but they are now moot, so we will not consider them.)

Consequently, Usenet hosts do not meet the necessary conditions for the system-caching safe harbor. (Incidentally, Usenet hosts with archival and search capability (like Google Groups and Guba, which come up later) assuredly do not.)

### **0.3.2 Transitory communications may protect traffic**

The only safe-harbor that might protect Usenet servers is the transitory-communications safe harbor, which is examined below.

(a) Transitory Digital Network Communications. <sup>-28</sup>

A service provider shall not be liable for monetary relief, or, except as provided in subsection (j), for injunctive or other equitable relief, for infringement of copyright by reason of the provider's transmitting, routing, or providing connections for, material through a system or network controlled or operated by or for the service provider, or by reason of the intermediate and transient storage of that material in the course of such transmitting, routing, or providing connections, if

-

The key phrase is the limitation of liability for “the intermediate and transient storage of that material in the course of such transmitting, routing, or providing connections.”

This is the highest hurdle: do Usenet transmissions, which arrive at the service provider and sit on its servers awaiting download, qualify? Let's consider it word-by word.

Is the ISP's “storage” of Usenet articles “intermediate”? Yes. It holds them as a service to users, waiting for users to pick them up.

Is the ISP's “storage” of Usenet articles “transient”?

From Webster's Revised Unabridged Dictionary (1913):

Transient \Tran"sient\, n.

That which remains but for a brief time. --Glanvill.

<sup>29</sup>

Yes. The archetypal ISP deletes articles to save space as new ones come in.

Do ISP store Usenet articles “in the course of such transmitting, routing, or providing connections”?

Indeed they do. The Network News Transfer Protocol's (Usenet's protocol, hereafter NNTP) is a set of instructions for *transmitting* and *routing* news.<sup>30</sup> By tradition, running a Usenet host means setting up a server that

---

<sup>28</sup><http://www4.law.cornell.edu/uscode/17/512.html#512.a>

<sup>29</sup>Thanks to emacs's M-x dict.

<sup>30</sup>Skeptics may shout “Hey! You're twisting the statute! This was obviously written to protect low-level TCP/IP routing, not Usenet routing!” After discussion with Jacob Glass and Professor Ed Felten, I realized that (a) Congress probably intended to shield ISPs

downloads articles from elsewhere and holds them for end-users, as RFC 977<sup>31</sup> (Internet protocols are described in documents called RFCs, or Requests For Comments) indicates:

### 1.3. Central Storage of News

For clusters of hosts connected together by fast local area networks (such as Ethernet), it makes even more sense to consolidate news distribution onto one (or a very few) hosts, and to allow access to these news articles using a server and client model. Subscribers may then request only the articles they wish to see, without having to wastefully duplicate the storage of a copy of each item on each host.

While the RFC says ‘Ethernet’, the logic could easily be extended to cable, DSL, or even dialup Internet connections. In 1986, when this RFC was written, dialup TCP/IP (as opposed to shell<sup>32</sup>) access, much less cable or DSL TCP/IP access, was unheard of. Internet access meant dialing into a shell account on a networked machine (like the one described in the RFC), or typing at the console of one of those networked machines at an Internet-connected facility. The basic idea of keeping a central ‘spool’ of articles and permitting users to read from it is the same, whether the users have separate accounts on the same multiuser machine, or separate machines connected to a local-area-network (or ISP), with the right to access a central news server. ISP caching of news is an integral part of the NNTP routing process, and thus meets this condition of the transitory-communications safe-harbor.

---

from liability for email transmitted through their servers, and (b) an email message also follows this arrive-and-wait pattern. It is hardly a stretch to surmise that this safe-harbor applies to high-level routing, as well. (Email may also be eligible for the user-initiated storage safe harbor. That *may* mean that its storage isn’t covered by the transitory-communications safe harbor. Determining a conclusive answer would entail studying the committee hearings on the subject, a digression which we must unfortunately omit due to constraints on space and time.)

<sup>31</sup><http://www.faqs.org/rfcs/rfc977.html>

<sup>32</sup>The difference between TCP/IP access and shell access is that TCP/IP access connects one’s computer directly to the Internet, letting one use web browsers, email clients, FTP clients, and the like, whereas shell access just provides a point-to-point connection to another computer that itself has Internet access. One interacts with this computer through a terminal window.

(1) the transmission of the material was initiated by or at the direction of a person other than the service provider;<sup>33</sup>

RFC 977 implies that there are two ways by which news arrives on an ISP's server: either a neighboring server asks it if the server wants more news and sends it if the server affirms (push), or the server asks its neighbors if they have new news (pull).

“A host desiring new news, or which has new news to send, will typically contact one or more of its neighbors using NNTP.”<sup>34</sup>

The first case, where news is pushed, clearly meets this stipulation. The second case, where news is pulled, requires a bit more analysis. An analogy may help.

U.S. Mail is not delivered directly to undergraduates at Princeton. Rather, in a nod to the pre-Benjamin Franklin days of the postal service, undergraduates must come to mail distribution centers to pick it up. There are six on the campus: one in each residential college, and one for upperclassmen in the Frist Campus Center.

The action of *picking up* the material must be initiated by the student, but the *transmission* of the material that appears in the student's mailbox is initiated entirely by other parties. Likewise, if a server has a news-pickup rule (e.g., “every day at 3 a.m., ask the neighbors if they have any new news”), the number and content of the articles it downloads is determined by someone *other* than the service provider. The transmission of a given Usenet article is initiated by its author. The fact that some servers use a pull (rather than a push) method to receive it is merely an artifact of the mechanics of Usenet, and changes nothing about the principles of the transmission process. Functionally, this stipulation distinguishes between receiving a letter in the mail and checking out a book from the library. News is most emphatically closer to the former. Just because one has to visit the post office to pick up mail from a P.O. box does not mean that one “initiated” the transmission of that mail to the box. Likewise, a mechanistic news-checking process, even if pull-based, still relies on other people to initiate transmissions.

---

<sup>33</sup>Note that this differs from the related stipulation regarding system caching. In system caching, the statute requires two parties: a transmitting user and a receiving user, where the receiving user initiates the transmission. Here, either user can initiate the transmission.)

<sup>34</sup><http://www.faqs.org/rfcs/rfc977.html>

Regardless of its use of push technology or pull technology, Usenet servers satisfy the stipulation that the transmission of news is initiated by the poster, not by the service provider.<sup>35</sup>

(2) the transmission, routing, provision of connections, or storage is carried out through an automatic technical process without selection of the material by the service provider;

Few ISPs actually read and censor Usenet content.

However, most ISPs select which groups they wish to carry. Some may exclude entire categories (this author's high school's network excluded the `alt.` hierarchy, for instance). Others might exclude all binary groups for space reasons, or certain groups like `alt.binaries.pictures.erotica.child` for obvious legal reasons. (This may expose the ISP to vicarious-infringement liability, because it is an admission that it could, for instance, filter out `alt.binaries.mp3` for copyright reasons.)

Is this safe harbor closed to ISPs that pick the newsgroups they want to offer? At one extreme, a judge might interpret the "automatic technical process" stipulation as sanctioning any automated process. This means that, so long as the service provider does not have a human censor, an computer-based filtering (keyword filtering, for instance) is permissible. Others might view the selection of groups as permissible, but not the use of software to filter articles within groups. Still others might demand that no filtration or selection take place between the ISP and the upstream provider.

In practice, requiring that ISPs engage in no traffic-filtering at all quite debilitating. Many ISPs refuse email messages that originate from domains known to harbor spammers<sup>36</sup> or from domains with software configured such

---

<sup>35</sup>Still skeptical? Do you think that "at the direction of a person other than the service provider" rules out pull *as well as* push news services, because the ISP has to make arrangements (of its own volition) to receive pushed news? (That is to say, the day to day operations may be automated, but the act is fundamentally a "pull"?) Don't be skeptical. Recall that while the news *service* is requested by the ISP (just as an email account is requested by its user), the actual transmissions of data (articles in this case, emails in the email case) are initiated by others — namely, their authors. If Usenet traffic were transmitted as are mailing lists, the articles would be sent directly to every subscribed user at every server as "push" transmissions. It is for efficiency's sake that Usenet uses this model, in which it appears that the transactions are initiated by the ISP, but they really begin with the authors of the articles.

<sup>36</sup><http://mail-abuse.org/rbl/>

that it permits spam redirection<sup>37</sup>. To claim that an ISP cannot dock in the transitory-communications safe harbor because they employ automated traffic filters may be a valid textual reading, but it means that the safe-harbor is essentially inaccessible for most ISPs.<sup>38</sup> We can conclude that *if* an ISP downloads anything less than the full newsfeed available from their neighbors (which is effectively all of the news on the Internet, if their neighbor is a commercial provider like Supernews), then they *might* not qualify for this safe harbor.

On the other hand, if an ISP's news server always downloads all news available on its neighbors, then it will easily satisfy this provision.<sup>39</sup>

(3) the service provider does not select the recipients of the material except as an automatic response to the request of another person;

This is easily met. ISPs do not pick the destination user for news articles – they hold them for all of their news users, as implicitly requested by the sender. (Such is the default behavior of Usenet.)<sup>40</sup>

(4) no copy of the material made by the service provider in the course of such intermediate or transient storage is maintained on the system or network in a manner ordinarily accessible to anyone other than anticipated recipients, and no such copy is maintained on the system or network in a manner ordinarily accessible to such anticipated recipients for a longer period than is reasonably necessary for the transmission, routing, or provision of connections; and

The first provision requires that the copies be made inaccessible to all persons except the “anticipated recipients.” This is easily satisfied: since news

---

<sup>37</sup><http://www.ordb.org/>

<sup>38</sup>Incidentally, mail-blacklists are an interesting hybrid case: while they are generally applied mechanistically (automatically) at the ISP, the selection of which sites actually go on the blacklist is usually human-assisted.

<sup>39</sup>A *reductio ad absurdum* into which I will not delve: if one is obligated to get all available news from one's neighbors, is one still permitted to pick one's neighbors? (In other words, may one choose between Supernews and a different upstream provider.) For that, too, is indirect content selection.

<sup>40</sup>This may mean that an ISP must grant all Usenet users access to all articles, because to restrict access would imply that the ISP was selecting the recipients of an article, whereas the article's author implicitly intended it to be reachable by everyone on Usenet.

is intended to be publicly accessible, then “anticipated recipients” includes everybody.

The second provision requires that content be discarded after a “reasonably necessary” time period has elapsed. This is also easily satisfied. News routing conventions specify that news is held for a given, finite period, and then discarded. (See the discussion of transience above.)

(5) the material is transmitted through the system or network without modification of its content.

Assuming that we consider the headers of a message to be distinct from its content (if not, then no email system would ever be protected by this provision, for email servers stamp messages that they route with their name and the date the message passed through), this is standard behavior.<sup>41</sup>

In sum, so long as an ISP has structured their Usenet service properly, they should be able to claim a spot in the transitory-communications safe-harbor, thereby protecting themselves from copyright-infringement lawsuits arising from content saved on their Usenet server.

### 0.3.3 Legal defense outside the safe harbor

Even if an ISP is denied docking privileges in one of the safe-harbors provided by the DMCA, it can still challenge a charge of contributory, vicarious, or direct infringement.

#### Beating contributory infringement claims

A claim of contributory infringement<sup>42</sup> requires three elements: one, that an act of direct infringement took place, two, that the accused was aware of the infringement, and three, that the accused did something to assist it.<sup>43</sup>

Unless you happen to be in China<sup>44</sup>, a commercial ISP will make it a point to be unaware of what you do (check their privacy policy) unless you force them to know what you’re doing. They will not read your news logs,

---

<sup>41</sup>Guba and Google Groups do not satisfy this provision: Guba decodes binary articles, and Google Groups reformats the messages and presents them in HTML.

<sup>42</sup>See <http://www.crblaw.com/faqs/copyrt.html#4> for more details on vicarious and contributory infringement.

<sup>43</sup><http://www.chillingeffects.org/piracy/faq.cgi#QID270>, also COS 491 seminar.

<sup>44</sup>Or Saudi Arabia, or any one of a handful of other illiberal nations.

so they have no idea what sort of infringement you might perpetuate on Usenet. Consequently, even if a user uses an ISP's Usenet feed to commit infringement by downloading copyrighted material, it is highly unlikely that the ISP could be held responsible for contributory infringement.

A truly gluttonous user might trip an alarm inside an ISP if they leech an inordinate amount of data, and the most likely place for a user to do that is the infringement-infested binary groups. A smart ISP, however, will write their privacy policy and their software in such a way that if they have to alert a user of excessive downloading, the ISP's employees are prohibited from knowing what the user was downloading (ISPs are advised to justify this as a privacy measure, rather than as blatant avoidance of contributory infringement liability).

The average ISP, consequently, will be easily able to avoid liability for contributory infringement claims.

### **Tiptoeing around vicarious infringement claims**

A claim of vicarious infringement may be served against an ISP whose service is employed in direct infringement if the ISP profits from the infringement, and the provider has the "right and ability to control users"<sup>45</sup> (presumably in such a way as to stop the infringement).

Except to the extent that users buy an ISP subscription for the sake of getting at copyrighted material on Usenet,<sup>46</sup> ISPs do not directly profit from infringement. In fact, the presence of infringing materials on a server, if it is discovered and sparks a lawsuit, can be rather costly to the ISP. However, as Professor Felten indicated and the von Lohmann paper affirms) the courts have interpreted "direct financial benefit" very broadly. For instance, the fact that Napster's facilitation of infringement lured more users to their service was held to be a direct financial benefit from infringement by the court in Napster, for it made their company "more attractive to investors."<sup>47</sup> Likewise, if any users sign up for an ISP because of its high-bandwidth, low-regulation News servers, with which they plan to infringe, then the ISP has profited from infringement.

There is little that an ISP can do to effectively filter infringing Usenet

---

<sup>45</sup><http://www.chillingeffects.org/piracy/faq.cgi#QID269>

<sup>46</sup>Users are much more likely to have infringement in mind with a commercial Usenet-only provider like Guba.

<sup>47</sup>[http://www.eff.org/IP/P2P/20010227\\_p2p\\_copyright\\_white\\_paper.html](http://www.eff.org/IP/P2P/20010227_p2p_copyright_white_paper.html)

content (part of the “right and ability to control”<sup>48</sup>). There are no machine-readable blacklists of content with which an ISP can filter their newsfeed, and ISPs can hardly be expected to construct their own (there is simply too much copyrighted content).

However, the mere fact that ISPs do not have an effective means of controlling *content* does not mean they lack an effective means of controlling *users*. The von Lohmann paper mentions that the courts have held that the “right and ability to control” criterion is satisfied by the ability to control users: in other words, “ability to terminate user accounts or block user access.” It is the rare ISP that lacks that ability.<sup>49</sup>

How ISPs are expected to prevent (not stop, but prevent) infringement with that ability alone is an open question.

Ironically, an ISP that tried to build its own blacklist (perhaps from the *Billboard* charts) might well find itself on the wrong side of a vicarious infringement suit from a copyright-holder whose works were omitted from the blacklist. A reasonable judge is unlikely to look kindly on such a case, especially if the ISP offers to add the plaintiff to the blacklist, but it exemplifies the trouble with these half-measures. Strictly speaking, the use of such a blacklist is *prima facie* evidence that the ISP thinks itself fully capable of controlling both its users’ accounts *and* filtering content they can access.

Can an ISP plausibly avoid vicarious liability by simply throwing up their hands at the sheer volume of Usenet articles? Judicial interpretation of the vicarious liability statutes may obligate ISPs to take reasonable measures to stop Usenet infringement, which may well mean making blacklists, even if only as a stopgap measure. Note that high-profile vicarious liability cases stemming from Usenet are rare, perhaps because (a) Usenet has a low profile, compared to modern P2P services, and (b) most ISPs probably comply with the DMCA and have gained the transitory-communications safe-harbor.

### **A digression on vicarious infringement**

The “there’s simply too much content for me to be expected to filter it, so despite my control over the servers, I actually *lack* the ability to control what my users are doing” defense against vicarious liability (which I will call the

---

<sup>48</sup>[http://www.eff.org/IP/P2P/20010227\\_p2p\\_copyright\\_white\\_paper.html](http://www.eff.org/IP/P2P/20010227_p2p_copyright_white_paper.html)

<sup>49</sup>[http://www.eff.org/IP/P2P/20010227\\_p2p\\_copyright\\_white\\_paper.html](http://www.eff.org/IP/P2P/20010227_p2p_copyright_white_paper.html)

*firehose*<sup>50</sup> defense) ushers in an interesting alternative scheme.

Usenet has addressed its spam problem with a system called NoCeM.<sup>51</sup> A good netizen, when she sees spam on a newsgroup, can post a digitally-signed NoCeM message to the `alt.nocem.misc` newsgroup. Other Usenetters, if they trust the poster's opinion, will not see the spam if they use NoCeM-enabled software.

If the RIAA and MPAA and the Pornography Industry Association of America were to start a NoCeM service (call it **NoPrC**<sup>52</sup>) that flagged infringing Usenet articles (perhaps by checking them against audio/video fingerprint database), could ISPs essentially be forced to use it?<sup>53</sup>

First, a new judicial interpretation of the vicarious liability doctrine would have to reject the firehose defense. Then the transitory-communications safe-harbor would have to be closed to Usenet traffic. With no DMCA defense, and a likely prosecution for vicarious infringement if they were caught with infringing material on their news servers, ISPs would be forced to either shut down their Usenet servers, or consent to filtering their newsfeeds with NoPrC. The RIAA, MPAA, and PIAA could sweeten the deal by offering vicarious-infringement immunity to ISPs that use NoPrC.

NoPrC would render the DMCA-infringement-notification process unnecessary on Usenet: to obliterate an article, the content-owner simply would tell the NoPrC operator to add the message-ID of a given article, *et voila* — it vanishes from most of Usenet. Might the RIAA and friends also use this to suppress Usenet discussions they don't like? Perhaps discussion of DeCSS? It's unlikely (if they want to get ISPs to consent to it), but not impossible.

If the triumvirate of content-owners first built up a precedent in court that denies safe-harbor to most Usenet servers, then offered immunity from infringement lawsuits to Usenet ISPs who used the system, only the most radical Usenet providers would refuse. (Others might simply drop Usenet altogether.)<sup>54</sup>

---

<sup>50</sup>A nod to the "getting an education here is like drinking from a firehose" quip recited by MIT inmates. (<http://www-tech.mit.edu/V105/N57/hose.57n.html>)

<sup>51</sup><http://www.cm.org/faq.html>

<sup>52</sup>"No Piracy"

<sup>53</sup>An ISP could apply NoCeM directly to its news-spool, but if they did, they might run afoul of the "don't select content" provision of the transitory-communications safe harbor. Most, by consequence, let users apply it themselves. While NoPrC would also violate that provision, the ISP would presumably only consent to use NoPrC in exchange for immunity from infringement suits from the biggest content-holders.

<sup>54</sup>Similar mechanisms (blacklists of file names, fingerprints, or other metadata) could be

## Eyeballing direct infringement claims

ISPs who are unaware of direct infringement by their users (those that are aware of it and take no action have little shelter under any law) are exempt from *criminal* copyright liability as assigned by the NET (“No Electronic Theft”) Act.<sup>55</sup> The NET act criminalizes even non-commercial violations of copyright, but stipulates that those infringements must be “willful.” The archetypal Usenet server is a robot, requiring no human intervention except when it breaks. Even if it plays the host to copyrighted articles, unless the ISP administration inspects its news cache, their role in the infringement is secondary: the direct infringers are those who post and download copyrighted articles. Its owners can hardly be described as “willful” infringers, especially if they have a policy of removing copyrighted material that comes to their attention (which a prudent ISP would have, to gain DMCA protection).

Can a Usenet-host also avoid civil liability for direct infringement (stemming from hosted infringing content) by demonstrating a lack of intent? It appears not: the law assigns liability to “[a]nyone who violates any of the exclusive rights of the copyright owner.”<sup>56</sup> The fact that the criminal code stipulated that violations must be “willful” implies that there do exist non-volitional copyright violations.<sup>57</sup>

The courts have invented<sup>58</sup> the doctrine of vicarious infringement to assign liability to parties who had no knowledge of an infringement, and did not participate directly in it.<sup>59</sup>

It remains possible to construct what computer scientists call a “pathological case” — a test that exploits a system’s worst flaws. For our purposes,

---

used for P2P networks: if the RIAA, et al. were to grant infringement immunity to ISPs who filtered out certain files from Gnutella and eDonkey traffic. (Fortunately for KaZaA users, ISPs would have less of an incentive to use this system, for KaZaA traffic is well falls well within the confines of the transitory-communications safe-harbor.)

<sup>55</sup><http://www.usdoj.gov/criminal/cybercrime/17-18red.htm>—

<sup>56</sup><http://www4.law.cornell.edu/uscode/17/501.html>

<sup>57</sup>Fred von Lohmann’s paper lends weight to this theory: “[I]f you make or distribute any copies (even if only in RAM) of copyrighted works, you may be held liable as a direct infringer. The court will not be interested in ‘control’ or ‘knowledge’ or ‘financial benefit’ or ‘material contribution.’ If you made or transmitted copies, you’re probably liable for infringement.” ([http://www.eff.org/IP/P2P/20010227\\_p2p\\_copyright\\_white\\_paper.html](http://www.eff.org/IP/P2P/20010227_p2p_copyright_white_paper.html))

<sup>58</sup><http://www.crblaw.com/faqs/copyrt.html#4>

<sup>59</sup>For details on the NET act and its history, see <http://stealth.kirenet.com/aleinss/piracy.pdf> <http://www.virtualschool.edu/mon/Outlaws/LaMacchiaIndictme> <http://www.eff.org/Legal/Cases/LaMacchia/>

that case is a hypothetical service provider called *Blackbeard*. *Blackbeard* is a free-web-hosting provider like geocities.com, but with two major differences: *Blackbeard* is not-for-profit, and *Blackbeard* makes it a point to be unaware of the content that its users upload – they never browse the site or look at the logs (“We’re volunteers,” they claim, “We have jobs. We’re too busy to play ‘policeman.’”). Consequently, *Blackbeard* became a haven for warez<sup>60</sup> distribution. *Blackbeard* responds to cease-and-desist letters directed at individual acts of infringement, but makes no other effort to track down and stop infringing users.<sup>61</sup>

*Blackbeard* does not make a profit, so vicarious infringement doesn’t apply. Its admins has no knowledge of what their site is being used to do except when it receives a cease-and-desist letter (and in that case, they remove the offending content), so contributory infringement doesn’t apply. To muddy the waters a little, let’s suppose that most of the content hosted on *Blackbeard* is noninfringing — but there are enough warez sites interspersed to be a problem.

*Blackbeard* is not liable under the vicarious doctrine. It’s not liable under the contributory doctrine. And if direct infringement liability requires willful infringement, then *Blackbeard* is not directly infringing. Is *Blackbeard* then permitted to escape liability entirely? Intuitively, *Blackbeard* should be liable for the infringement perpetuated with its system. Does this mean that the “direct = willful” assertion is incorrect?

If it is incorrect, we can conclude that every service provider is liable for any content hosted on its servers — even if the provider didn’t put it there. ISPs are immunized against liability for content posted by their users if they comply with the DMCA. But such a protection is not afforded to newsfeeds – which are posted automatically by the ISP. The articles in the newsfeeds are themselves written by other users, and transmitted to the ISP, but a skeptic might balk at calling this “storage at the direction of a user,”<sup>62</sup> which the statute requires.<sup>63</sup>

---

<sup>60</sup>Internet slang for illegal copies of software programs.

<sup>61</sup>This bears a minor resemblance to the LaMacchia case, which prompted the passage of the NET act. <http://www.virtualschool.edu/mon/Outlaws/LaMacchiaIndictment.html>

<sup>62</sup><http://www4.law.cornell.edu/uscode/17/512.html#512.c>

<sup>63</sup>Even if the skeptic were persuaded, the user-content safe harbor requires that the ISP not receive a “financial benefit directly” from the infringing material. As von Lohmann indicated in regards the Napster case, the fact that infringing content attracts new users is enough to satisfy the benefit test attached to vicarious infringement. (The language of

Despite this analysis, the direct-infringement issue may be moot. Given that the “firehose defense” against vicarious liability is not terribly persuasive, ISPs will be vicariously liable for infringing material regardless of how the direct-infringement issue is resolved. ISPs provide news servers to attract users (to boost profits), and if an ISP’s news server attracts users who use it to infringe copyrights, even if the ISP would rather they did not, they will be subject to vicarious liability.

### 0.3.4 How to run a Usenet server without losing a lawsuit

The first lesson to draw from this section is to stay within the DMCA’s safe harbors. Outside of them, ISPs will find their rickety legal-vessels buffeted by a typhoon of infringement lawsuits.

To stay within the safe harbor, an ISP is advised to have Usenet news pushed to their servers, rather than pulled in by them — it’s a minor change, but it means that the plaintiffs must show “the transmission of the material was” *not* “initiated by or at the direction of a person other than the service provider.” We demonstrated earlier that, but for the eccentricities of the NNTP protocol, transmissions are initiated by Usenet posters — but this strategy raises one more hoop for the plaintiffs to jump through. ISPs are also advised to perform no filtering whatsoever (this bolsters the “firehose defense” against vicarious infringement claims — the more articles hitting the server, the more difficult it is to drink from the firehose).

Finally, to preserve their lack-of-knowledge defense against contributory infringement claims, ISPs should avoid looking for infringing material, and delete any that is reported to them.<sup>64</sup> Abstinence from filtering also improves the lack-of-knowledge defense — the less the ISP staff sees of the newsfeeds, the better.

---

this statute is similar to the language of the doctrine, so it is likely that the same precedent would apply.) Few Usenet ISPs would be able to demonstrate conclusively that no-one had ever signed up for their service to download infringing binaries.

<sup>64</sup>Fred von Lohmann advised P2P service authors on how to minimize their liability exposure; much of his advice applies here. (Usenet is itself a peer-to-peer system.) [http://www.eff.org/IP/P2P/20010227\\_p2p\\_copyright\\_white\\_paper.html](http://www.eff.org/IP/P2P/20010227_p2p_copyright_white_paper.html)

## 0.4 What Hath Google Wrought?

Google Groups provides a permanent archive of the articles posted to Usenet since 1981. They exclude binary files, articles with the `X-No-Archive` header, and articles suppressed by order of the author or by legal action.<sup>65</sup>

We will draw upon the analysis in the last section and define Google in terms of the average ISP's Usenet server. Google Groups does everything an average ISP's Usenet server does, but it has a few *additional* attributes that narrow its range of legal defenses.

### 0.4.1 $\Delta$ s: No safe harbor for Google Groups

Google Groups loses the transitory-communications defense because it never allows articles to expire. On a higher level, the real difference is that Google does not resemble the usual model of Usenet news routing as discussed in RFC 977, so its actions do not qualify as customary routing.

Google Groups incorporates the capabilities of a search-engine, a technology which is in itself eligible for DMCA protection. However, that protection (if Google Groups is eligible for it) only extends to the “search” component of their site.

#### (d) Information Location Tools. -<sup>66</sup>

A service provider shall not be liable for monetary relief, or, except as provided in subsection (j), for injunctive or other equitable relief, for infringement of copyright by reason of the provider referring or linking users to an online location containing infringing material or infringing activity, by using information location tools, including a directory, index, reference, pointer, or hypertext link, if the service provider -

Even if Google Groups's search engine meets the DMCA's requirements (I have omitted them), the search engine defense only immunizes it against liability for its links. It does not protect against liability for permanent archiving. Consequently, if charged with infringement, Google must actually go toe-to-toe with the plaintiffs — it can't simply steam into a safe harbor and drop anchor. Consequently, we will consider its defenses against infringement claims.

---

<sup>65</sup><http://groups.google.com/googlegroups/help.html>

<sup>66</sup><http://www4.law.cornell.edu/uscode/17/512.html#512.d>

### 0.4.2 $\Delta d$ : Changes to direct infringement liability

Google’s direct-infringement defenses do not differ significantly from those of an ordinary Usenet host. Google went to more of an effort<sup>67</sup> to pull in old Usenet posts, but unless it can be demonstrated that they willfully included infringing material in their archive, their criminal liability for direct infringement is nil, and their civil liability remains constant.

### 0.4.3 $\Delta c$ : Changes to contributory infringement liability

While at least *some* of the articles on Google Groups contain infringing material, and while Google Groups undeniably helps people locate and download infringing text posted to Usenet, an accuser would be hard-pressed to demonstrate that Google was actually aware of any of those infringements, nor could Google reasonably be expected to trawl their database for them. The reported infringements are often quite esoteric.<sup>68</sup> Short of doing a search for the text of every copyrighted work in existence, there’s no way that Google could find them all.

Google remains safe from contributory-infringement claims, unless they start ignoring DMCA infringement notifications (known colloquially as “take-down” letters), or otherwise fail to act against infringing material that is brought to their attention.

### 0.4.4 $\Delta v$ : Changes to vicarious infringement liability

Google sells advertising on the Google Groups site, so it profits from all use of the site, infringing or not. It also has the “right” to control users, and, by consequence, stop infringement: it operates both the index and the archive, and has the power to remove any items it pleases.

Possession of the “right” to control its users does not mean that Google has the “ability” to stop all forms of infringement before it hears complaints. Google’s archive contains 700 *million*<sup>69</sup> articles. Google can, and does, act on specific cases of infringement, but it can’t be expected to find them on

---

<sup>67</sup><http://groups.google.com/groups?selm=51f8d6cf.0108281758.7af9e804%40posting.google.com>  
[http://web.archive.org/web/20011020213925/http://groups.google.com/googlegroups/archive\\_hunt.html](http://web.archive.org/web/20011020213925/http://groups.google.com/googlegroups/archive_hunt.html)

<sup>68</sup><http://www.chillingeffects.org/notice.cgi>

<sup>69</sup><http://groups.google.com/googlegroups/help.html#about>

its own. (Note that this is the firehose defense, which may or may not pass judicial muster.)

Strictly speaking, Google *does* profit from infringement, and it *can* control what its users do. But given Google's record, it's unlikely that a claim of vicarious infringement would prevail: it does not profit from infringing articles any more than it does from non-infringing articles (except to the extent that some infringing articles might draw more users to the site), its business model does not leverage the infringing material to lure users to the site, Google follows<sup>70</sup> the designated-agent specification requirements specified<sup>71</sup> in the DMCA, and it has a record of responding<sup>72</sup> to takedown notices. A plaintiff who brought a complaint without first beseeching Google to remove the offending material with a DMCA takedown notice would probably face tremendous skepticism.

#### 0.4.5 Advice for Google: how to avoid lawsuits

To enjoy the benefits of a cache of all of Usenet since 1981, Google had to leave the DMCA's transitory-communications safe-harbor. Nevertheless, they remain safe from most infringement suits. Their presumed lack of knowledge of infringement nullifies contributory infringement claims. The fact that their business does not substantially benefit from infringement (unlike Napster, for instance, Google profits primarily from users looking for legitimate material) protects them from most vicarious-infringement claims (as does the firehose defense). Hosting the articles on their web server does expose them to direct-infringement liability, but given their history of removing content in response to complaints, and the fact that they do not archive binary content (which is often infringing), it is unlikely that they will lose a lawsuit over it. Their model is the best that it can be, given that they had to leave the safe harbor to make the archive available.

### 0.5 The Guba Question

Guba.com, like Google Groups, archives Usenet content. Guba.com, unlike Google Groups, archives binary content. More specifically, Guba archives

---

<sup>70</sup><http://www.google.com/dmca.html>

<sup>71</sup><http://www4.law.cornell.edu/uscode/17/512.html#512.c.2>

<sup>72</sup><http://www.wired.com/news/politics/0,1283,51233,00.html>

and indexes only images and videos.

The protections afforded to Usenet server operators shrink when applied to Google Groups. They shrink further when applied to Usenet binary archives, primarily because those archives have fewer non-infringing uses. As with Google Groups, Guba cannot dock at one of the DMCA's safe harbors. But unlike Google Groups, Guba will have a more difficult time fighting off allegations of vicarious and contributory infringement, much less direct infringement. This is a consequence of the nature of online multimedia content, which tends to not be produced by the people who post it, and is more likely to be infringing.

### 0.5.1 $\Delta d$ : Changes to direct infringement liability

As above, unless Guba knowingly and willingly hosts infringing material, they have not committed criminal direct infringement. They may be vulnerable to civil infringement suits, but can always take the Google defense — that they respond rapidly to copyright-infringement suits.

Guba will have a tougher time proving the “knowingly and willingly” assertion, let alone the rapid-response-to-discovery-of-infringing-material assertion: three clicks from the front page of their site (go to “<http://www.guba.com>”, click “Video,” click “Comedy,” click “`alt.binaries.multimedia.sitcoms`” are links to episodes of “Will and Grace” and “Malcolm in the Middle.”<sup>73</sup>

This raises a question: could it be that Guba is *knowingly* and *willingly* hosting copyrighted material using it for its business, making a show of accepting DMCA complaint notifications, but ignoring the fact that nearly all of their content is infringing material for which no one has yet bothered to submit a complaint?

If that is the case, then, intuitively, Guba is committing *willful* direct infringement. If a direct-infringement lawsuit were launched, a subpoena might determine whether this sort of “wholesale infringement” is part of the Guba business model.

---

<sup>73</sup><http://www.guba.com/c/412/v/582/1.phtml>

## 0.5.2 $\Delta c$ : Changes to contributory infringement liability

Google Groups may have been able to assert that not only was nearly all of its content noninfringing, but that the tiny bit that was infringing (a) did not come to their knowledge unless someone filed a complaint, in which case they removed it, and (b) hardly offset the substantial noninfringing uses of their site.

Guba has a far weaker claim of plausible deniability: first, their content has a far higher likelihood of being infringing, and second, it's trivially easy to find obviously infringing content on their website ("Malcolm in the Middle" in three clicks, for instance).

Guba also faces a contributory infringement issue that an ordinary Usenet provider does not face: Guba repackages binary articles to make them easier to download. Typically, large binaries are posted to Usenet in several dozen chunks, to get around message-size constraints. To extract the content, the user must find, assemble, and decode all of the articles. There exist end-user software packages<sup>74</sup> that automate this process, but Guba has done it *for* the users. Essentially, it does not provide a Usenet service, or even an interface to Usenet, but rather, files skimmed from Usenet.<sup>75</sup>

While the greatest source of infringements on text-based newsgroups probably consists in people copying and pasting articles from Yahoo!, CNN, Reuters and the like (which very, very rarely gains the attention of the news agencies), most of the content on text-based newsgroups is conversation, commentary, or original work.

Image and video groups, on the other hand, see a far lower ratio of legitimate-to-illegitimate use, simply because creating videos is considerably more difficult than typing messages. Image creation is today trivial, but few people are sufficiently uninhibited (or have sufficiently uninhibited friends) as to sustain the 60,000 files-per-day (much of which is presumably pornog-

---

<sup>74</sup>Debian, a Linux-based operating system project that maintains a large software repository at [packages.debian.org](http://packages.debian.org) has four packages that can turn any newsfeed into your personal Guba: `ubh` at <http://packages.debian.org/stable/net/ubh.html>, `brag` at <http://packages.debian.org/testing/news/brag.html>, `aub` at <http://packages.debian.org/unstable/news/aub.html>, and `pimppa` at <http://packages.debian.org/unstable/gnome/pimppa.html>.

<sup>75</sup>Google Groups does some touch-up work, to: it displays the messages as HTML, and sorts them by date and subject for easy browsing..

raphy, judging by the proportion of links to pornography to links to other material on Guba's front page) that Guba claims to retrieve from Usenet — while only a content breakdown of Usenet could demonstrate this with certainty, Occam's Razor dictates that most pornography is probably *not* produced by its copyright-agnostic exhibitionists.

Ironically, it may be Usenet's most reviled demographic that provides the majority of the legitimate multimedia content: porn spammers. The majority of Guba's front-page links are to adult content. Many adult sites, in trolling<sup>76</sup> for potential customers, might post pornographic multimedia to Usenet, emblazoned with their logos and covered with links to their web sites. Adult sites have certainly tried this with modern P2P systems:

“We love file trading,” said Kevin Blatt, sales director for the Triple X Cash, which operates the Collegefuckfest.com and Rectalrooter.com websites. “Why? It's called greed. We've found a way to monetize that sharing.”

Blatt's company embeds hidden links in video clips and sends the short movies out on the sharing networks. Then, when a file-swapper downloads a clip and clicks somewhere in the video's frame, he's taken to one of Triple X Cash's sites. The company gets 25 to 40 “joins” — \$30 monthly subscriptions — per day from this technique, according to Blatt.<sup>77</sup>

Even though spammers are not known for their honesty, they presumably spam Usenet with images from their own sites (unless they're *really* dishonest), to which they presumably have rights (spammers are like cockroaches — they can scurry. A copyright-violating commercial website, on the other hand, is a sitting target.)

### 0.5.3 $\Delta v$ : Changes to vicarious infringement liability

Most ISPs run Usenet servers as a perk — one more feature to encourage users to sign up. By contrast, Guba, like Google Groups, has its Usenet service at the heart of its business model: it profits purely from people choosing their Usenet services over others'. Vicarious infringement requires that the service provider in question profits from infringement. Guba.com will have

---

<sup>76</sup>Trolling in the old sense of the word, not the new sense.

<sup>77</sup><http://www.wired.com/news/business/0,1367,58665,00.html> (hat-tip: pho-list)

a far more difficult time than did Google Groups in demonstrating that it does not profit from infringement – in fact, that may not be possible. The sheer proportion (and of placement of porn-folder links may well indicate that most of Guba’s content is porn, which indicates that much of it is probably infringing.

Guba’s firehose defense is also narrowed because they’re discarding a good deal of traffic (namely text traffic). Nevertheless, they have a substantial amount of content (if their figures are to be believed — cross-posts and reposts may inflate these numbers significantly):

“GUBA decodes 60,000 images and videos per day from USENET using automated software. That’s one image or video every 1.44 seconds, 24 hours a day, 365 days per year. For obvious reasons, we cannot and do not review or editorialize this material.”<sup>78</sup>

If a substantial number of those images are infringing, a judge might determine that it is necessary for Guba preview that material to weed out the infringing content.

#### **0.5.4 How could Guba shield itself from lawsuits?**

Guba’s best course of action is to recast themselves as something closer to Google Groups. If they, for instance, included *all* binary content (rather than just images and videos), they would have a better shot at casting themselves as a legitimate search service (this author knows of no sites that archive every binary article posted to Usenet.) If Guba indexed text content as well, they could bill themselves as a comprehensive Usenet archive — a title to which not even the great Google can lay claim. Either of these approaches would make them look substantially less like a site devoted to infringing copyrights. (Revising the front page so that it’s no longer possible to find copyrighted material in three clicks might help, too.)

All of those approaches would expand Guba’s noninfringing uses, and create plausible deniability about instances of infringement (which bolsters the contributory-infringement defense). The addition of more content also bolsters the anti-vicarious-infringement firehose defense.

It is not the mere expansion of Guba’s offerings that reduces its liability, it is the fact that expansion engenders a change in the character of the

---

<sup>78</sup><http://www.guba.com/static/aboutguba.phtml>

service. If Guba becomes more general, its utility rises, its legitimacy rises, and any allegations that it is devoted to wholesale copyright infringement would wane. Some of its users might continue to engage in infringement, but, like the malefactors who use P2P networks, they need not spoil the fun for the legitimate users.<sup>79</sup>

## 0.6 Wrap-up: Infringement defenses from Usenet to Google to Guba

Over these pages, we've watched the broad defenses hoisted by ISP-based Usenet servers taper into the thinner shields raised by Google Groups, and into the gossamer veil cast over Guba.com. Changing the technical details of a site can alter its liability exposure substantially, but there exist two primary rules for fending off copyright lawsuits stemming from Usenet:

**More articles are better.  
Less intervention is better.**

The more articles you keep on your site, the better. More articles increase your site's utility — and a general, useful tool is harder to paint as a device of infringement than a narrow one.<sup>80</sup> Keeping more articles on the site also strengthens the weak “firehose defense” against vicarious liability, and underscores the lack-of-knowledge defense against contributory liability. If it would be impossible for a few human beings to vet all of the articles that come through the your newsfeed in a 24-hour period, then you can't be expected to monitor it. (There is one exception to the “more is better” rule: if you're running an ordinary Usenet server, be sure to expire the articles in a reasonable amount of time. If you do not, their storage is no longer “transient,” and you lose the transient-communications safe harbor.)

---

<sup>79</sup>Or so concluded Judge Stephen Wilson when he threw out the RIAA's case against the companies that made the Grokster and Morpheus P2P clients: [http://eff.org/IP/P2P/MGM\\_v\\_Grokster/030425\\_order\\_on\\_motions.pdf](http://eff.org/IP/P2P/MGM_v_Grokster/030425_order_on_motions.pdf)

<sup>80</sup>General tools can be used for evil. A savvy user can find infringing MP3s with Google, but no-one sues them for it. See <http://golem.ph.utexas.edu/distler/blog/archives/000139.html> (hat-tip: [http://www.freedom-to-tinker.com/archives/2003\\_04.html](http://www.freedom-to-tinker.com/archives/2003_04.html))

Less intervention means fewer employees examining the news service. Not only does establishing the news-service as a low-maintenance operation reduce the credibility of a plaintiff who insists that you have an “ability to control,” it also means that the ISP’s employees are less likely to see copyright infringements, which bolsters the lack-of-knowledge defense against contributory liability.

If you ever plan to run a Usenet service, take this to heed. Be general. Don’t focus on images and video, for instance. Go for volume. Pull in as much as possible, without filtering it. And don’t work too hard. Confine your interaction with the news server to maintenance, as much as possible. Keep these in mind, and your vulnerability to lawsuits won’t disappear, but your opponents’ chances of winning them will plummet.

81

---

<sup>81</sup> L<sup>A</sup>T<sub>E</sub>X help: <http://cyberbuzz.gatech.edu/sga/grad/latexstyle.html>