Re-engineering Usenet to Provide Censorship Resistant Publishing

Richard Clayton[1], George Danezis[1], Emil Sit[2],
Christopher T. Lesniewski-Laas[2], M. Frans Kaashoek[2], Robert Morris[2]
[1] University of Cambridge Computer Laboratory
[2] MIT Laboratory for Computer Science

Abstract

Usenet is a highly successful peer-to-peer publishing system and has proved to be good at resisting the suppression of material. However, it has failed to cope with recent attacks that have used the legal system. We present a scheme based on a large distributed cache that improves on Usenet's general censorship resistance. When the default mechanisms appear inadequate, authors can deploy robust publishing methods and article storage mechanisms in an entirely seamless manner.

1 Introduction

One of the most interesting results of the communications revolution centred around the Internet has been the transformation of journalism – in its widest sense as the reporting of and commenting upon contemporary events. It is no longer necessary to own a hugely expensive printing press and complex distribution system to promulgate one's opinions. Internet access is invariably bundled with web space and with direct access to discussion forums such as Usenet. The recent popularity of web 'blogs' shows that there is a very widespread interest in presenting individual views of communal data.

As traditional publishers embrace these Internet-based paradigms, the focus of their business is moving away from the mechanics of placing information in front of people, towards a rôle as a filtering system, selecting information to emphasise and discarding the rest. With this shift comes an increasing interest in 'collaborative' or 'distributed' filtering where this editorial rôle is performed by collectives or by the readers themselves. There is considerable interest in systems that support decision making, enforce editorial policies, and permit divergent views of the data stream.

The initial, libertarian, view from the 1980's and 1990's that the Internet would provide forums where people could say anything they wished has proved to be incorrect. High profile court cases on defamation, such as *Godfrey v Demon Internet* [16], and the removal of web pages, such as the republished *Radikal* articles [23], are just the tip of an iceberg, with ISPs regularly acting as 'judge and jury' and suppressing material even though a court might not have done so [19]. This occurs because the administrative entities serving up the material are susceptible to legal attack and have very little direct interest in maintaining its availability.

These legal attacks have now moved on to attempting to suppress not only the source material but also links to where it might be found. The courts have had considerable difficulty in this area, handing down a patchwork of contradictory decisions in different jurisdictions [15][12][13]. However, the general point remains that the reason these attacks succeed is that the publishing function is centralised and a change to more distributed systems would make attacks more costly to mount – if not entirely impractical.

When censorship has become public there has been broad-based resistance to it. High profile cases about websites have invariably led to ad hoc replication of the content in other jurisdictions, such as occurred with the JET report [9]. There are now signs of a legislative rethink, with the Law Commission's second report on defamation [19] drawing attention to the damage being done to guarantees of free speech. Their concern comes from the propensity of British ISPs to remove material that is defamatory even when there might be a legal defence of justification ('it's true'). In several recent cases such as the PopBitch story about David Beckham [3] and allegations of rape by soccer players [4], lawyers have contacted ISPs who have then suppressed what is little more than ill-informed bar-room gossip.

So there is a spectrum of censorship occurring, from the often banal in Western democracies through to the most heavy-handed suppression of anti-government commentary in less liberal states [22]. Historically, Usenet has coped well with technical attempts to suppress material, so it is a good candidate for a censorship resistant publishing system. However, in Section 2 we show how it is regularly failing to resist legal attack. Therefore, in Section 3, we set out a new high-level design to address issues of anonymity and resistance to the removal of material. The scheme is not designed to prevent machine operators from obeying the law but nevertheless we intend for articles to continue to be served by making it very inconvenient for any legal challenge to succeed in censoring material, short of shutting down the system altogether. Finally, in Section 4 we show how this design enables us to provide not only traditional Usenet semantics, but also systems for the distributed publishing and syndication of website hosted material.

2 Usenet

Usenet is a mature, successful, peer-to-peer discussion forum and publishing system. An article posted to a particular Usenet node is distributed via a flood-fill protocol to all other participating nodes. Individual nodes can refuse to carry particular articles, but cannot affect their overall distribution and availability. However, authors are vulnerable to attacks that trace articles back to their source and reading is prevented wherever all available nodes are required to operate suppression policies.

As is often the case, the sex industry is at the forefront of technical and legal developments, though other groups who seek to suppress material may not be far behind. Several publishers of adult websites have sought to protect visitor revenues by ensuring that their images do not circulate on Usenet. Legal action has been taken by Color Climax [8], Suze Randall [2], Playboy [21] and particularly ALS Scan, whose case against RemarQ [1] is an oft-cited precedent. These activities, which now take the form of cease-and-desist orders served upon individually identified newsgroup posters, have considerably reduced the open posting of particular images.

In the United Kingdom the problem of illegal images of children on Usenet is dealt with by the Internet Watch Foundation (IWF) who run a reporting hotline for the public. Originally the IWF just asked ISPs to remove illegal articles. More recently they have banned several newsgroups because their names suggested the presence of illegal material, even though articles within them might in fact have been lawful.

2.1 Reader and Poster Anonymity

Usenet provides good receiver anonymity because the flood-fill protocol gives no information as to where an article might be read. Further, most sites keep no logs of reader activity, partly because reading patterns have long been perceived to be sensitive data and partly because the logs would be extremely large and would serve no operational purpose. However, should a particular 'sysadmin' become interested in the local readership of a particular newsgroup or article then this is, in principle, trivial for them to determine.

Anonymity can be improved by using a 'trusted intermediary' that guarantees not to log. Also, in recent years, Deja News (later acquired by Google) has provided universally available web-based access to text articles, and there are numerous anonymising web proxies. Alternatively, the news server might be accessed via an anonymising TCP service such as Tor [14]. However, issues of latency and efficiency come to the fore and the practical deployment of industrial strength systems remains a significant challenge.

Material can be suppressed by making people too scared to post articles because they fear being identified as the author. Anonymous connections and trusted intermediaries can address this as well. More powerful techniques such as using remailer networks [11] to access 'mail2news' gateways can also be used, though a determined attacker could conceivably succeed in closing the small number of injection points currently operating.

2.2 The Problem of 'Spam'

One can deliver censorship resistance by making deletion of stored material impossible. However, if at the same time one removes the ability to prevent particular individuals from posting then this threatens the entire viability of Usenet. The reason for this is 'spam'!

Spam has been a problem on Usenet since the first high profile events in 1994 [24][6]. By the end of the decade bulk-posted material formed a third of the news feed and robotically generated 'cancels' formed another third – a cancel can only remove a single article. Spam remains a problem, but it is now only 5% of the feed. This may be partly due to the spammers migrating to email, but is in no small part due to the widespread usage of automatic filtering systems like CleanFeed [20]. Recently spammers have been 'proxy hijacking', so that abusive material is posted by assuming the identity of an innocent user. Nonetheless, the ability to trace the posters of spam and deny them further service remains a key factor in preventing abuse.

We can therefore portray the search for censorship resistance for Usenet as being in no small part a search for effective anti-spam techniques that cannot be used to censor other material.

3 High-level Design

We now present a brief overview of our new design for Usenet and explain how it improves censorship resistance. As shown in Fig. 1 the author A writes an article

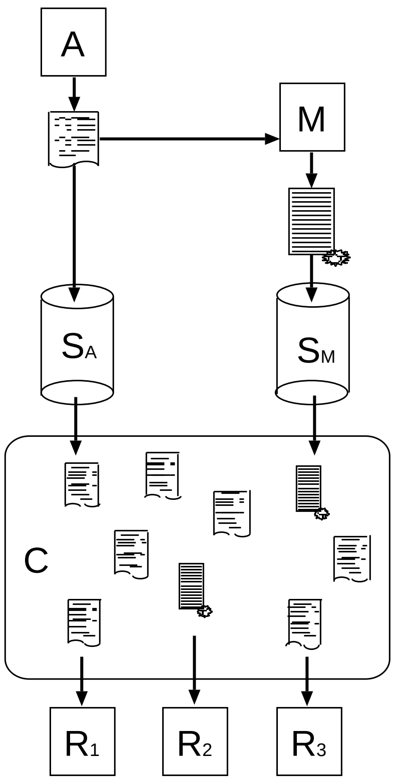

Figure 1: Proposed new high level design for Usenet

and places it into a storage system S_A. They request a moderator M to announce that the article exists and to report its location. This request may, optionally, be pseudonymous via a remailer. The moderator periodically places a signed list of articles of which they approve, as an article itself, within their own storage system S_M. Readers R_i learn of moderated lists via some out-of-band channel, such as an external web page or an article on another list. They fetch the lists and use the location information to retrieve the articles themselves. A communal caching system C holds a working set of lists and articles, which are stored under self-authenticating keys so that it is impossible to serve up fake material undetected. We provide no explicit protection against denial-of-service attacks, but the caching system will ameliorate this issue.

Although this system appears to be concentrating on protecting the articles, in fact the challenge is to protect the articles containing the moderators' lists whilst providing a reasonably efficient scheme for accessing their contents. We are always aiming to ensure that an attacker working within the legal system will find it necessary to resort to wholesale closure of the system because nothing short of that will be effective.

3.1 Article Storage

The communal cache will be a large distributed storage system such as Chord-DHash [10]. Articles are split into blocks, replicated using an n of m erasure code, and stored on remote machines participating in the Chord ring. An adversary who wished to suppress an article would therefore need to remove more than $m-n$ blocks from the system by interacting with many machine operators, possibly in different jurisdictions.

However, this is not the only threat to article storage, since machine operators may voluntarily choose to suppress particular articles and, since the system's storage capacity is not unlimited, they will have to 'expire' older articles on a simple FIFO basis at each particular Chord node. The authors therefore need to preserve their articles within a storage system of 'last resort' that can be accessed when the article is not available within the communal cache. The moderators' lists will provide the information necessary to locate this storage.

Where the author is concerned that attempts will be made to censor their material, they will wish to put in considerable effort to ensure that it remains available. In particular they might publish a 'location' for it which is merely an access handle for a specifically censorship resistant system such as a chain of Chaumian mixes [7]. It is a key element of our design that the censorship resistance of individual articles no longer arises solely because the material has been 'spread randomly' but because the author is now able to place the material where the incentives to resist removal are consistent with the author's security goals. Since legal attacks are often slow, defenders will have a little time to assess how best to preserve material that is being attacked.

3.2 Lists of Articles

In specifying a moderator rôle we do not mean to imply that a human must perform it. An automatic system that signs a list of every submitted article would be sufficient. Although our system preserves Usenet's model of organising articles into newsgroups we have not provided a hierarchical namespace. In our scheme, the public key for the moderator's signature becomes the newsgroup name. This makes the arbitrary suppression of names because of a perceived link with content impossible and removes any opportunity for central control over which names are considered to be valid.

The readers select the moderators that they wish to use by methods that are outside our scheme (they may consult search engines or accept recommendations from friends), but ultimately they will be basing their decision upon the reputation that accretes to each signing key (and associated pseudonym). It should be carefully noted that articles can appear upon an indefinite number of lists. We view this multiplicity of views as being extremely desirable not only because it permits users to select a reliable stream of on-topic material, but because it allows 'syndication' of articles into collections that were not envisaged by the original author.

The scheme is resistant to censorship by category, the suppression of entire newsgroups. Although it may be possible to remove a particular list of articles – albeit the lack of semantic content may be legally significant in some jurisdictions – it is trivial to republish under a new signing key, which will be discovered by the same out-of-band methods as the original. The moderators' lists are public, though there is some 'security through obscurity', which will last until the first insider aware of the significance of the signing key turns informer.

4 Flexibility

4.1 Replicating Existing Usenet Semantics

As already indicated, existing Usenet methods for submitting articles can continue to operate. Robot gateways can publish lists of articles and provide a 'last resort' storage system on behalf of any author. Also, those who like the current hierarchical Usenet newsgroup scheme are free to collect an ordered list of signing keys ('newsgroups') from an authority figure such as 'the Usenet cabal' and even hold ballots to decide whether further keys should be added to this master list.

Continuing with the same paradigms does of course perpetuate spam, which is seen to be effective by its senders because an advertising message is distributed 'for free' and will be put in front of many individual readers. We believe that our system can be configured to provide several disincentives to spammers. Filtering submitted articles at the moderation stage will prevent the material getting onto lists in the first place. We also expect existing 'spam cancellers' to continue their activities. However, instead of increasing the system load by issuing cancel messages, they need only provide signed lists that are a subset of those that other moderators (likely robotic moderators) have issued. This is not dissimilar to the NoCeM scheme which has been operating since 1995 [5].

We remain ambivalent as to whether there is value in providing automated methods of removing spam articles from the communal cache. Although such a scheme can be used for the greater good, it also provides a method of instantaneous censorship of articles, which must then be re-fetched from the author's storage system. We intend to develop replacement algorithms that can depreference the storage of spam (which will not be widely read because it will not appear on moderator lists), concentrating instead in caching useful material.

4.2 Co-operative Web Sites

Existing journalistic web sites such as the loose amalgamation of 'Indymedia' sites [17] also work well in the proposed scheme. These sites report events, often those overlooked by mainstream coverage. They then permit individuals who have witnessed the event, or who just have opinions about it, to add their own comments (so far, the scheme is very similar to that operated by a site like Slashdot). Important stories are then 'syndicated' to other Indymedia sites elsewhere in the world, so that more people may view the same story on their own local website. Unfortunately, the existing mechanisms for syndication run into difficulties, particularly with handling comments, when the editorial standards at the destination site differ from those at the syndication source.

The Indymedia sites can make direct use of our new Usenet design. They would each maintain their own signing key and would syndicate articles by adding them to their own lists. An appropriate mix of automatic and manual procedures can then be used to deal with subsequent comments without, of necessity, affecting the content of other websites. The websites themselves merely present the appropriate parts of the new Usenet in an

HTML-based format and can be quickly moved to other servers in the event of legal attempts at censorship or physical denial-of-service attacks.

5 Conclusion

Usenet is the most successful peer-to-peer system yet developed. We have described a new design for Usenet, which is intended to make it more censorship resistant by making legal attacks more difficult. In particular we propose ensuring that those who wish to ensure that data is highly available in the face of censorship must take the responsibility of securing that data rather than just distributing it widely and hoping.

References

[1] ALS Scan Inc v RemarQ Communities Inc, Feb. 2001, No 00-1351 (CA-99-2594-JFM)

[2] AVN News: Suze Randall Hunting Cyberthieves, 29 Jun 2000. http://www.suze.net/preview/copyright/Cyberthieves.htm

[3] BBC News: Psst! Heard the latest? 2 Dec 2002. http://news.bbc.co.uk/1/hi/in_depth/sci_tech/2000/dot_life/2527005.stm

[4] BBC News: Football 'rape' case prompts net debate, 1 Oct 2003. http://news.bbc.co.uk/1/hi/uk/3156848.stm

[5] Cancelmoose: The NoCeM FAQ v0.93, 1998. http://www.cm.org/faq.html

[6] L. Canter: Green Card Lottery– Final One? 12 Apr 1994. http://groups.google.com/groups?selm=2odkr9$3r5@herald.indirect.com

[7] D. Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Comm. ACM 24(2), 1981, pp. 84–88.

[8] Color Climax Corporation, APS; Digital Graphic Systems, AB; and International Media Company, BV v "Bile", Feb 1999, 98Civ, 4956(ERK), Korman, J. Go, M.J.

[9] Cyber-Rights and Cyber-Liberties: Press Releases about the availability of the JET Report on the Internet, 4 Aug 1997. http://www.xs4all.nl/~yaman/jetrep.htm

[10] F. Dabek, M.F. Kaashoek, D. Karger, R. Morris and I. Stoica: Wide-area cooperative storage with CFS, Proc. 18th ACM Symposium on Operating Systems Principles (SOSP'01), Oct 2001.

[11] G. Danezis, R. Dingledine and N. Mathewson: Mixminion: Design of a type III anonymous remailer protocol. In 2003 IEEE Symposium on Security and Privacy, pp. 2–15, IEEE CS, May 2003.

[12] Michelle Delio: Deep Link Foes Get Another Win, Wired News, Jul 8 2002. http://www.wired.com/news/politics/0,1283,53697,00.html

[13] Michelle Delio: Deep Linking Takes Another Blow, Wired News, Jul 20 2002. http://www.wired.com/news/politics/0,1283,54083,00.html

[14] R. Dingledine et al: Tor: The Second-Generation Onion Router, 2003, to appear.

[15] Michelle Finley: Attention Editors: Deep Link Away, 30 Mar 2000. http://www.wired.com/news/politics/0,1283,35306,00.html

[16] Godfrey v Demon Internet Ltd, QBD, [1999] 4 All ER 342, [2000] 3 WLR 1020; [2001] QB 201

[17] Independent Media Center (Indymedia): http://www.indymedia.org

[18] Law Commission for England and Wales: Aspects of Defamation Procedure: A Scoping Study. CP Special 4, May 2002.

[19] Law Commission for England and Wales: Defamation and the Internet: A Preliminary Investigation: Scoping Study No 2. CP Special 5, Dec 2002.

[20] J. Nixon: Cleanfeed – Spam filter for Usenet news servers, 16 Aug 2002. http://www.exit109.com/~jeremy/news/cleanfeed/

[21] Playboy v Sanfilippo, March 1998, (S.D.Cal.1998)

[22] Privacy International and The GreenNet Educational Trust: Silenced: an international report on censorship and control of the Internet, Sep 2003. http://www.privacyinternational.org/survey/censorship

[23] Statewatch: Judge orders Indymedia to remove internet links. http://www.statewatch.org/news/2002/may/03dutch.htm

[24] C.L. Thomas IV: Global Alert For All: Jesus Is Coming Soon, 19 Jan 1994. http://groups.google.com/groups?selm=9401191510.AA18576@jse.stat.ncsu.edu

All URLs verified 9 Nov 2003.