In: *On the Identity Trail: Privacy, Anonymity and Identity in a Networked Society*, Eds. Ian Kerr, Carole Lucock and Valerie Steeves (Forthcoming, Oxford University Press)

TrackMeNot: Resisting Surveillance in Web Search[1]

DANIEL C. HOWE
Computer Science
New York University
and
HELEN NISSENBAUM
Media, Culture & Communication
New York University

TrackMeNot (TMN) is a Firefox browser extension designed to achieve privacy in web search by obfuscating users' queries within a stream of programatically-generated decoys. Since August 2006, when the initial version of TMN was made publicly available, free of charge, there have been over 350,000 downloads. TMN protects web users against data-profiling by simulating HTTP search requests to search engines with queries extracted from the web. In an attempt to mimic users' search behavior this basic functionality is augmented with several technical mechanisms: dynamic query-lists, real-time search awareness, live header maps, burst-mode queries, and cookie-anonymization. We describe each of these mechanisms, evaluate its strengths and weaknesses, and demonstrate how the consideration of values directly informed design and implementation. In the discussion section we conceptualize TMN within a broader class of software systems serving ethical, political and expressive ends. Finally we address why web search privacy is particularly important and why TMN's approach, for the present moment, is both legitimate and necessary.

---

1. INTRODUCTION

TrackMeNot (TMN) is a lightweight Firefox browser extension designed to achieve privacy in web search by obfuscating a user's actual searches amidst a stream of programmatically generated decoy searches. Since August 2006, when the first version of

[1] Many individuals and institutions have contributed in essential ways to this paper. For critical feedback on earlier versions of this paper, we thank audiences at the Haifa Center of Law and Technology Conference on Law of Search Engines (2006); the Conference on Computer Ethics: Philosophical Equiry (2007); the Annual Meeting of the American Association, Eastern Division (2007); the A2K2 Conference, Information Society Project, Yale University (2007), the Poynter Center, Indiana University, Bloomington (2007) and the Santa Fe Institute (2007). Additional thanks to Jinyang Li, Robb Bifano, the Mozilla foundation, MissingPixel™, and the NYU Media Research Lab. Thanks also to the reviewer for this volume, who guided several key improvements. We are indebted for help with TrackMeNot itself to enumerable users around the world who cheered us, critiqued us, and generously offered marvelous tips. We extend a special thanks to Michael Zimmer for all these contributions, and more. Support for this project came from the NSF award CCR-0331542: *Sensitive Information in a Wired World (*or PORTIA: Privacy, Obligations, Rights in Technologies of Information Assessment). The idea for

TMN was made publicly accessible, free of charge, there have been over 330,000 downloads[2].

In August 2005, public awareness of the ubiquitous practice of logging and analyzing users' web-search activities was raised when front page articles in the mainstream press revealed that the United States Department of Justice (DOJ) had issued a subpoena to Google for one week's worth of search query records (absent identifying information) and a random list of one million URLs from its Web index. These records were requested in order to bolster the Government's defense of the constitutionality of the Child Online Protection Act (COPA) then under challenge. When Google refused the DOJ's initial request, the DOJ filed a motion in a Federal District Court to force compliance. In March 2006, swayed by Google's arguments that the request imposed an unreasonable burden, would compromise trade secrets, undermine customers' trust in Google, and have a chilling effect on search activities, the Court granted a reduced version of the first motion, ordering Google to provide a random listing of 50,000 URLs and denied the second motion seeking the query records. One year later, however, the illusion that our web searches are a private affair was further pierced when a news investigation revealed that from anonymized search query logs provided to the research community, the identities of certain searchers had been extracted from personal information embedded in their search terms [Hansell 2006, Barbaro 2006]. Other media reports followed about how the major search companies (Yahoo!, AOL, MSN & Google) log, store and analyze individual search query logs.

Setting aside the details of these two highly publicized cases a few disquieting points remain: one, that search queries are systematically monitored, scrutinized, and indefinitely stored by search service providers; two, that for all we know, they are shared with third parties[3]; and three, that policies governing these practices are unilaterally set by search companies with little indication, or say, given to individuals about what gets done with their search records[4], scholars [24], and government agencies in the U.S. and beyond [FTC 2007]. Responding to this interest, search companies have offered several

---

TrackMeNot was hatched in a series of stimulating conversations with PORTIA colleagues at a project retreat.

[2] TrackMeNot may be downloaded from the website: http://mrl.nyu.edu/~dhowe/TrackMeNot/ and from the Mozilla add-on website: https://addons.mozilla.org/en-US/firefox/addon/3173.

[3] For example, court documents indicated that AOL, Yahoo!, and Microsoft had not been issued subpoenas because they had complied with the government's request.

compromises, none of which, with the possible exception of Ask.com, are adequate or fully transparent. We believe these policies and practices challenge foundational moral and political principles of our society.

In Western liberal democracies, freedom of expression and association are among a set of core values protected directly through laws (for example, the U. S. Constitution) and indirectly in the design of public institutions. Protection of liberties is also extended to activities considered supportive of these values, such as education, research, reading, and communication. As many of these activities have migrated online, so has the recognition that robust civil rights protections are required online as well. It has not required a great leap to compare the role of public libraries and town squares in promoting core freedoms with that of the Web, functioning as it does not only as a repository of information, but also as a public and personal medium for communication and association. Just as we expect freedom and autonomy in the former, "brick and mortar" versions, so we should in the latter, digital electronic versions. Information search and retrieval behaviors are part and parcel of these activities, profoundly reflecting who we are, what we care about, with whom we associate and how we live our lives. For behaviors that open a window to the personal and political commitments of individuals, existing practices and policies of search engine companies seemed clearly inadequate. Less clear, however, is how to pursue reforms to achieve necessary levels of protection, and who should or would lead the way.

Among potential agents of reform, the evident structure of incentives indicated that two with the greatest power to effect change – government, pursuing new laws and regulations, and search companies, revising internal policies – would be least likely to support such change. Intransigence and inaction in the face of early challenges bore this expectation out. For the first source of reform, government, search logs are an obvious and potentially important repository of information about individuals' interests and transactions, a valuable component of the vast stockpile of personal information assembled under the more lenient terms governing the collection and uses of information by the private sector [Birnhack 2003]. Actions that might constrain access to such information or limit its availability is not likely to be attractive.

---

[4] Since that time, interest in the issue of search privacy has greatly magnified, drawing attention from citizens, advocacy organizations.

With the second potential source of reform, search companies themselves, we predicted that they would be unlikely to welcome externally directed restraints on how their logs are treated and used. For a start, there is the general suspicion corporate actors hold for any imposition of third-party regulation. With their interests best served by as little oversight as possible, search companies attempt to mollify worried users and regulators by insisting that unconstrained access to and use of query data is an essential necessity for running their businesses, as, for example, explained by Eric Schmidt, CEO of Google "…the data helps us to improve services and prevent fraud."[Schmidt 2007][20] Although there is no reason to doubt this explanation, it masks a story that is never front and center in search companies' public rhetoric, but behind concerns of critics and privacy advocates, namely, the ways unconstrained assembly and use of detailed search query logs factor into the massive profit engine of personalized advertising.

A third source of reform is new government regulation or legislation steered by direct citizen action or the advocacy of privacy organizations such as the Electronic Privacy Information Center (http://www.epic.org), Privacy International (http://www.privacyinternational.org), the Center for Democracy and Technology (http://www.cdt.org), or the Electronic Frontier Foundation (http://www.eff.org). Although this approach has already born fruit, for example, the widely publicized report [Privacy International 2007], it will require an orchestrated effort of diverse parties, including many (government actors, search companies, advertisers, etc.) with a stake in maintaining unrestricted access to search logs. Although, ultimately, this is the most sound hope for lasting change, measurable success is most likely a long-term prospect only.

TrackMeNot represents a fourth alternative. Overcoming some of the obstacles inherent in the others, it *offers* control directly to those most motivated to seek reform, providing a relatively near-term even if imperfect solution. The hope, too, is that alternatives like TrackMeNot might bring reluctant parties into meaningful dialog about search privacy.

2. DESIGN CONSTRAINTS

> "The constraints of technique, resources, and economics u*nderdetermine* design outcomes. To account fully for a technical design one must examine the technical culture, social values, aesthetic ethos, and political agendas of the designers." [Pfaffenberger 1992]

Our approach to the development of TrackMeNot builds on prior work that has explicitly taken values into consideration in the software design [Freidman 2002, Flanagan 2005]. Throughout the planning, development and testing phases, we have integrated values-oriented concerns as first-order 'constraints' in conjunction with more typical engineering concerns like efficiency, speed, and robustness. Specific instances of such values-oriented constraints include: transparency (in interface, function, code, and strategy), personal autonomy (users need not rely on 3rd-parties), social protection of privacy (distributed/community-oriented action), minimal resource-consumption (cognitive, bandwidth, client and server processing, etc.), and usability (size, configurability, ease-of-use, etc.) Enumerating such values-oriented constraints early in the design process enabled us to iteratively revisit and refine them in light of the specific technical decisions under consideration [Flanagan 2005]. Where relevant in the following section, we discuss ways in which TMN's technical mechanisms benefited from this values-oriented approach.

3. TECHNICAL MECHANISMS

TrackMeNot, written in Javascript, C++, and XUL, is a Firefox browser extension designed to hide users' web searches in a stream of decoy queries. Query-like phrases are harvested by TMN from the web and sent, via HTTP requests, to search-engines specified by the user. To augment this basic functionality and frustrate attempts by search engines to distinguish between actual and generated queries, a range of mechanisms were implemented to simulate users actual search behaviors more effectively. These mechanisms and the design constraints informing their implementations are described in the following sections.

3.1 Dynamic Query-Lists

To maintain control in the hands of users TMN operates solely on the 'client', accessing no servers or 3rd-party sites during its operation. To support this design constraint while still maintaining unique query lists for each instance of TMN, we employed a mechanism we called *dynamic query-lists*, which function as follows. Upon installation, each instance of TMN contains a seed list of query terms gathered from publicly available lists of popular recent search terms (see Fig. 1 for a sample from such a list).

```
fashion, tv guide, barbie, neopets, bit torrent, xbox, angelina jolie,
nintendo, jennifer lopez, jennifer aniston, local weather, anime,
jokes, recipes, music lyrics, games, iraq, global warming, north
korea, hillary clinton, barack obama, dick cheney, zodiac, music and
lyrics, bone cancer, lena katina, iran, canada, veronica mars, lost,
the constitution, valerie plame, karl rove, halliburton, Iceberg,
global warming, world map, earth day, southern cross, spiderman 3, 300
movie, borat, shrek, bill of rights, ghost rider, Hawaii, dubai,
mexico, freedom of speech, Chelsea, London, kurt vonnegut, shaha riza,
yuri Gagarin, knut, Virginia tech, wellness, copyright law, health,
yoga, fishing, golf, Israel, Syria, Iraq, Pakistan
```

Fig 1. Sample from a TMN seed list.

From these seed terms (several hundred per client), TMN issues its initial queries. As operation continues, individual queries from this set are randomly marked for substitution. When a marked query is sent, TMN intercepts the search engine's HTTP response and attempts (non-deterministically) to parse a suitable 'query-like' term from the HTML returned. If, according to a series of regular expressions tests, the substitution is successful, this new term replaces the original query in the query list and the substitution mark is removed. This new term is now a member of the current query list and included as a potential future substitution candidate. Over time, each client 'evolves' a unique set of query terms, based in part on the random selection of queries for substitution, in part on the non-deterministic query extraction from HTML responses, and in part by the continually changing nature of web search results (generally yielding different results for the same search on different days). Fig. 2 shows examples from the query list above (Fig. 1) after several weeks of TMN operation. With dynamic query lists, TMN is able to avoid the use of any central or shared (and necessarily trusted) repository of query terms while still frustrating the filtering schemes to which a static list is vulnerable.

```
Turning carbon dioxide into fuel, Online Student Services, free
essential software, business globalization solutions, National
Pasta Association, Share your life with friends, Demand Financial
Suite, este calitatea produselor, Chicago Symphony Orchestra, This
film contains violence, Expects below Average, Emergency Contact,
bodies have been established, residential real estate, American
Heritage Month, Manhattan Athletic Club, healthcare support
occupations, people cannot realize their dreams, green chemistry
breakthroughs, Free online versions, Also find tools, Hope Press
```

Fig 2. Sample from an 'evolving' query list.

3.2 Real-time Search Awareness

Real-time search awareness (RTSA) is a second mechanism developed to improve TMN's capacity to mimic searchers' actual behavior. As TMN evolved, it became clear that it would need to 'know', in real-time, when users had initiated a search at one of the engines selected by the user. To facilitate this, the RTSA module examined each outgoing request from the browser and, via a series of regular expressions unique to each search engine, alerted TMN when the user was initiating a search. This feature, proved increasingly important, enabling the development of several other mechanisms (described below), which required knowledge of the user's current behavior, whether initiating a search, performing a series of searches, or engaged in other non-search activities.

3.3 Live Header Maps

Initially, development efforts focused on simulating the behavior of searchers in general. In later versions however, several features were introduced that enabled TMN to adapt to the behavior of specific users. Whereas the *TMN Control-Panel* (described below) allows users to manually configure TMN to more closely mimic their own search behavior, *live header maps* operate automatically to adapt TMN-generated queries to specific data sent by the client browser. This data generally varies according to browser version and operating system, as well as the search habits of specific users. To facilitate this adaptive behavior, TMN maintains a set of variables (per search engine) representing the header fields and URLs for the search most recently issued by the browser (see Fig. 3). These dynamically updating variables allow TMN to reproduce, in its own requests, the exact set of headers the browser has last used.

```
Url -> http://www.google.com/search?hl=en&client=firefox-
a&rls=org.mozilla%3Aen-US%3Aofficial&hs=nxM&q=hello&btnG=Search
User-Agent -> Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US;
rv:1.8.1.11) Gecko/20071127 Firefox/2.0.0.11
Accept -> text/xml, application/xml, application/xhtml+xml,text/html;
q=0.9, text/plain; q=0.8, image/png, */*; q=0.5
Accept-Language -> en-us,en; q=0.5
Accept-Encoding -> gzip, deflate
Accept-Charset -> ISO-8859-1, utf-8; q=0.7, *; q=0.7
Keep-Alive -> 300
Connection -> keep-alive
Referer -> http://collection.eliterature.org/1/
```

Fig. 3. An example header map for a Google web search.

Similarly, the specific URL last used to access a search engine is also maintained, so that, for example, if one user searches via the Google toolbar and another via the Google home page, TMN requests will mimic the header values for each. The RTSA module facilitates this functionality by only allowing updates to these variables when the user is initiating a new search at one of their selected engines.

3.4 Burst-Mode Queries

Another functionality enabled by RTSA was termed *burst-mode querying*. In initial versions of the software, semi-random intervals were used to temporally space TMN requests, with the average of these intervals set by the user. To more closely mimic actual user behavior, burst-mode triggers a batch of queries within close proximity to an actual user search (as detected by RTSA). By using this mode in conjunction with randomized intervals, users could 'blend' the two behaviors, employing more or less of each as desired. Further, by limiting bandwidth and processing use (for both client and search-engine) while dynamically adjusting to the use-patterns of the individual user, burst-mode operation allowed TMN to meet another design constraint, namely, lowered client (and network) resource-use.

3.5 Cookie Anonymization

A final mechanism enabled by RSTA is TMN's handling of search-related cookies[5]. When a TMN user enables the *cookie-anonymization* mechanism, cookies sent by the browser to the search engine are intercepted and stored by TMN. User-testing showed this strategy to be preferable to simply blocking or repeatedly deleting search engine cookies, as many users desired access to other functions of the search provider (e.g., GoogleMail for a Google searcher) which require the cookie. Thus, if a cookie is part of a user's search request, or its subsequent response, it is blocked by TMN (and in the request case, stored). Conversely, when a cookie-bearing HTTP request is *not* destined for a selected search engines (again monitored by RSTA), the cookie is ignored by TMN.

[5] Cookies have been employed by search engines as a means of aggregating data on a particular user over time, even when the user has not logged-in. In recent years, cookies have become somewhat notorious as a method by which advertising services, most famously, DoubleClick™, are able to compile *cross-site* profiles.

Perhaps counter-intuitively however, TMN does forward users' real cookies along with its own generated queries so that if a cookie-dependent profile exists for the user (linking requests from multiple IP addresses for example,) the TMN-generated data will be added to the profile, while the user's real searches will not. For an intuitive verification of this behavior, users have performed Google searches after enabling Google's *Search-History* function (which maintains a user-specific search history). With TMN's c*ookie-anonymization* module enabled, they find that TMN-generated queries appear in the history, while their actual queries do not.

3.6 TMN Control Panel

A range of user-configurable parameters allow users to further customize TMN's behavior (see Fig. 4). These include options to enable/disable TMN itself, the status bar display, query-bursting, and cookie management. Additionally users may select which search engines they wish to target, select an average query-frequency, and manage TMN's logging options. Finally, the *Control Panel* features buttons enabling users to view the current query list and action-logs within the browser itself, and to access the TMN website for additional information.

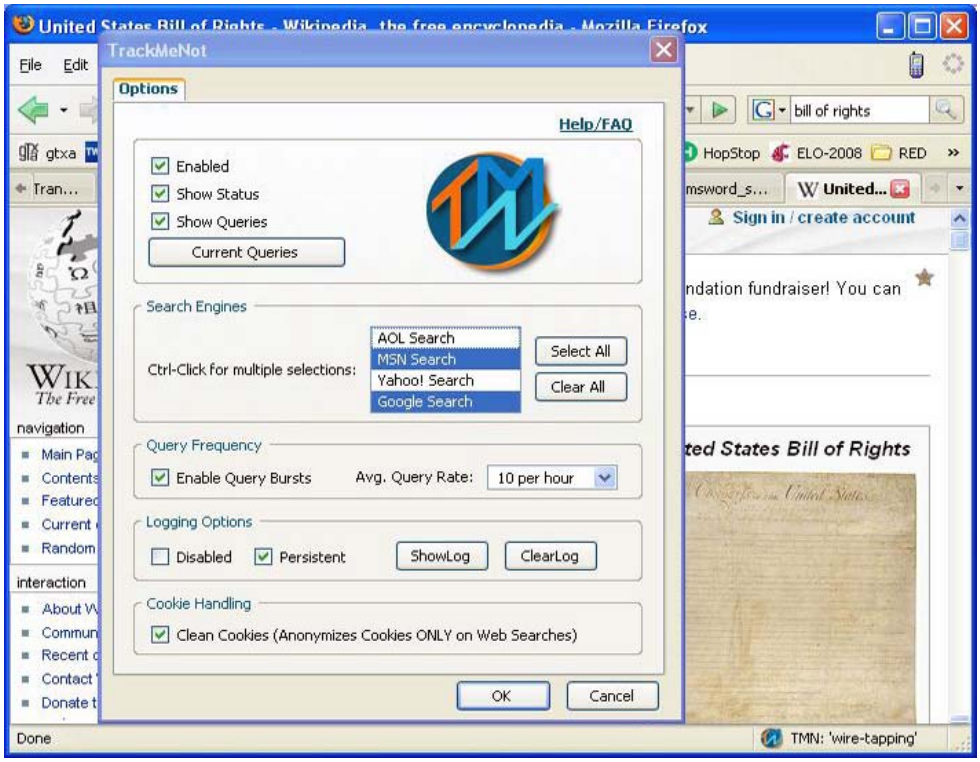

Fig. 4. The TrackMeNot Control Panel

Providing direct and real-time access to system operation (logs and query-lists), was directly informed by the design constraint of transparency. Further, TMN was released as free and open-source under a Creative Commons License [CC 2007] with all source files included (in plain-text format) in every download, allowing technically sophisticated users to examine the inner workings of the code and verify that it functioned as described. Similarly, our intentions and the specific technical decisions made to achieve them were published in straightforward, non-technical language on the TMN website (accessible directly from within the software itself).

4. EVALUATION: STRENGTHS AND WEAKNESSES

Evaluation of TMN was conducted iteratively throughout development and relied on solicited and unsolicited feedback from a range of groups including users, developers, software reviewers at Mozilla (where TMN is co-hosted), and a range of privacy and security advocates. While the question of whether TMN 'worked' seemed at first a simple one, we soon noticed how users' behaviors, goals, expectations, and perceived risks

shifted the meaning of the question. Analysis of feedback was thus often a two-part process, first determining the users' orientation, then examining the users' feedback in light of their respective goals, concerns, etc. Through such analysis we were able to identify at least 4 distinct groups of users, though individuals often identified with more than one of these.

One group was interested in TMN's ability to cloak searchers' identity and thus prevent any and all search activity from being traced back to them. We recognized that there were at least four mechanisms through which searches could be identified: 1) identifying information included in search queries (name, zip, phone, social security number, etc.); 2) static IP addresses linking searches across sessions; 3) explicit login to search engines (often for mail or other services); and 4) persistent cookies linking any of the above to users' search activities.

Although various prototype versions of TMN had included code to generate arbitrary personal information to mask actual identifying information, this strategy was not energetically pursued. This was largely because TMN was not designed to mask IP addresses and thus could not prevent identification via the IP addresses logged by search engines with every query, nor those maintained by users' ISPs. We pointed users interested in such tools to various proxy-based solutions [Tor 2007], linked from the TMN FAQ). Contrary to the assertions of some critics[6], TMN was not presented as a light-weight replacement for proxy-style solutions, but rather as very different approach (with a distinctive set of strengths and weaknesses.) To begin, proxies generally require users to grant some degree of trust to a 3rd party, whether a centralized server, or some 'exit node' representing the last hop in a 'distributed' solution. In the past, such exit nodes have been abused for a variety of purposes[7], or simply blocked by those not wishing to

[6] http://www.schneier.com/blog/archives/2006/08/trackmenot_1.html

[7] In September 2007, Dan Egerstad, a Swedish security consultant, revealed that he had intercepted usernames and passwords for a large number of email accounts, by operating and monitoring Tor exit nodes. On November 15, 2007, he was arrested on charges stemming from discovering and publishing this information. As Tor does not, and by design cannot, encrypt the traffic between an exit node and the target server, any exit node is in a position to capture any traffic which is not encrypted at the application layer, e.g. by SSL. While this does not inherently violate the anonymity of the source, it affords added opportunities for data interception by self-selected third parties, greatly increasing the risk of exposure of sensitive data by users who are careless or who mistake Tor's anonymity for security. [From http://en.wikipedia.org/wiki/Tor_(anonymity_network)]

receive their traffic (Google and Wikipedia being prime examples here[8]). While specific TMN users could also easily be blocked, once identified by a search-engine (at least for the duration of their IP address), it would take a very different kind of effort to block all such users. With Tor, for example, the identification of a single proxy node could result in the blocking of many thousands of user requests.

While a full discussion of the relative strengths and weaknesses of proxy-based solutions (including problems with internationalization and potential vulnerabilities to traffic analysis attacks[9]) is beyond the scope of this paper, it is worth noting the comparative 'user-friendliness' of such solutions in comparison with that of TMN. At least at the time of this writing, proxy-based solutions have been notoriously difficult for non-experts to set up, configure and use. They generally involve multiple components (e.g., a local executable and a browser plugin) which must be installed and configured to communicate correctly, at which point it is often still unclear what exactly the proxy is doing. This situation has been considered serious enough that privacy advocates (and third-party companies) have begun providing builds of popular software already containing say, a Tor configuration, to eliminate these difficulties for users[10]. This differs noticeably from TMN's one-click-and-restart installation and subsequent transparency of operation.

Of course there is no reason, as we have maintained in our FAQ, that these different approaches cannot be used together to additive effect. In fact, we believe this to be a rich area for future research. On the other hand there are difficulties faced by web-users that are equally troublesome for all proposed solutions. An example is the increasingly common case where a user wishes to search the web while being explicitly logged into a search engine, say for its free email services. Here, none of the proposed solutions, whether TMN, proxy servers, or others, offer much help.

---

[8] See http://www.boingboing.net/2006/09/07/google-blocking-priv.html, by Cory Doctorow discussing Google's blocking of Tor Nodes; also http://simple.wikipedia.org/wiki/Wikipedia:Bans_and_blocks regarding Wikipedia's policy to block all requests from anonymizing proxies (including Tor).

[9] See Murdoch, S. J. and Danezis, G. 2005. Low-Cost Traffic Analysis of Tor. In *Proceedings of the 2005 IEEE Symposium on Security and Privacy* (May 08 - 11, 2005). SP. IEEE Computer Society, Washington, DC, 183-195. DOI= http://dx.doi.org/10.1109/SP.2005.12

[10] See OperaTor, a preconfigured bundle including the Opera Browser, Tor, and Privoxy (http://archetwist.com/opera/operator) and the XeroBank Browser, a Firefox derivative with an integrated Tor configuration (http://xerobank.com/xB_browser.html).

Another group of users worried about being targeted due to 'hot-button' searches, that is, stigmatized or taboo subjects like 'anarchy', 'HIV', or 'drug-use'. To protect such users adequately, TMN would need to generate a range of similarly 'hot' query terms, an area on which we experimented through development, and that showed particular promise with the addition of dynamic queries in version 4.1x. Having found that dynamically evolved queries tend to stay within general topic areas, we reasoned that with hot terms in the initial seed list TMN would generate some number of extreme, or at very least, unwelcome surprises, e.g., potentially offensive or NSFW (not safe for work) queries which would be displayed publicly in the browser's status bar.

This worry was of particular concern to users in a third group, who wanted TMN to avoid hot button issues entirely, citing worries over social stigma, job loss, and even potential arrest. This group, preferring generic and innocuous noise, was primarily interested in TMN as a way to mask the true nature of their online searches in order to avoid wholesale aggregation and profiling, advertising and marketing purposes being the most salient. This clash between groups two and three is evident in the following two excerpts, respectively:

> "Some of them would have to have HIV, some of them would have to be contemplating suicide, some of them would have to be anarchists, etc. Maybe you wouldn't want to have pedophiles and Terrorists in the mix... or people growing hydroponic marijuana?" [anonymous user from group #2]

> "I downloaded and installed the plug-in you developed. I just turned it off when I noticed that the search term it had generated was 'free russian porn boys'. I'm a little confused. I understand the rationale of trackmenot, or I thought I did, but how does associating my ip with searches for gay porn fit in? If my employer logged this search, it could put my job at risk". [anonymous user from group #3]

The concerns of a fourth group of users stemmed from potential civil rights violations due to web-search monitoring. Though, like the third group, they also worried about the logging and aggregation of search queries for the purpose of profiling, they were interested in TMN mainly as a disruptive tool for protecting citizens against agents of government who might be engaging in various search surveillance and aggregation practices.

As we saw these groupings emerge and considered how they might guide the iterative process of feedback and development, it was clear (per the popular saying) that “you can’t please all the people all the time.” Accordingly it made sense to focus on what we believed to be TMN’s greatest strength, that is, providing protection against aggregation and profiling of individual search queries. This meant anticipating various

ways that TMN-generated queries might be detected and filtered, a process aided greatly by the helpful feedback of critics and enthusiasts alike volunteering their insights and pointing out potential weak links. We conjecture, in large part due to the many iterations of the software emerging from such discussions that considerable effort would now be required to 'defeat' TMN and successfully filter user queries from TMN queries. We further surmise that such filtering efforts would require significant resources and would still be likely to generate a number of false positives, that is, user queries mistakenly judged to be TMN generated.

The critiques we considered drew attention to various ways in which TMN queries could be different enough from user queries to inform an filtering algorithm to distinguish between the two. We have worked most recently to address versions of this critique based on 3 aspects of TMN's operation: query-timing, click-through behavior, and query-term analysis. Timing-based critiques have argued that the timing patterns of TMN's queries, even when randomized, were different enough from human-generated queries to be detectable. Our solution to this critique was to add burst-mode querying so that TMN queries can occur only when users are actually searching at a targeted engine.

'Click-through' critiques pointed out that TMN queries are never followed up with onward clicks to outgoing links on a search results page. In fact, we have developed test versions of TMN with this functionality, but have chosen not to release them until we have adequately understood the potential impact on existing advertising business models. We also concluded that while one might 'know' that queries leading to click-throughs were user-generated, inferring the converse, that queries yielding un-clicked results pages were TMN-generated, would surely result in significant false-positives (user-generated searches discarded along with those generated by TMN). Missed user queries of this type might be particularly costly to search companies aiming to improve their performance through personalized query log analysis.

A final version of the filtering argument focused on the nature of the query terms themselves, stating that these were not 'real' enough to fool a sophisticated learning algorithm with access to the vast amounts of data that search engines have already collected. Although dynamic query generation from actual web-pages has clear virtues, it is difficult to state how effective this strategy would be if search engines were willing to allocate significant resources to overcoming it. It is possible that a machine-learning algorithm, focusing on query content, perhaps in conjunction with other factors, could be trained to identify a high percentage of TMN users, possibly even a high percentage of

specific queries themselves. Obstacles to defeating TMN are primarily the costs of human and material resources (engineers, hardware, software), costs of false-positives (discarded user queries), potential costs to reputation (users and/or media outcry), and the potentially increasing maintenance costs required to handle past and future versions of TMN with different behaviors. Although the extent of such costs are difficult to assess, especially in light of the vast resources available to search companies, we hope they might be high enough to make other collaborative, trust-oriented compromises seem more attractive.

Unfortunately, we have little insight into counter-measures that may be taken by targeted search engines. Unlike tactics such as URL-format changes or IP-address blocking which will be readily apparent, other possibilities, such as the filtering strategies discussed above, might occur unnoticed. We have thus far benefited from the 'many eyes' of the developer and open-source communities, prodding us to consider such counter-measures, as well as from the evaluations of users and critics.

5. POLITICS THROUGH TECHNOLOGY

Conceiving of technologies as forms of political action builds on an intellectual tradition that includes figures such as Langdon Winner and Bruno Latour, who have argued that technical devices and systems may embody political and moral qualities. Lawrence Lessig, and others [Friedman 2002, Introna 2000] have explored these ideas in the context of information technologies and digital networks. Allied with this academic tradition, though not necessarily in direct dialog with it, activist designers, software developers and digital artists have leveraged the malleability of IT and the openness of network protocols to develop utilities that are expressive of particular political commitments or mediate transactions in politically charged ways[11].

Placing control in the hands of users, which we adopted as one of TMN's design constraints, is not all that makes it political. Its political character comes from the way it enters into and attempts to reshape a particular aspect of individuals' relationships with social actors far more powerful than themselves on nearly every measurable dimension --

[11] Examples include: GNU (http://www.gnu.org/); Creative Commons tools, (http://creativecommons.org/); P3P (http://www.w3.org/P3P/); Adrian Ward's AutoIllustrator (http://www.medienkunstnetz.de/works/autoillustrator/); Wikipedia, (http://www.wikipedia.org/); and the Radical Software Group's Carnivore (http://r-s-g.org/carnivore/).

wealth, mastery over technology, and access to power. TMN, by allowing individuals to set limits of the flow of personal information, belongs in a class of technical tools that serve as amplifiers of social resistance or political voice. Relying on neither the largesse, nor the permission of others, especially those with potentially clashing interests, TMN provides, for some users, a means of expression, like a political placard or a petition. For others, provides a practical means of resistance in a vein similar to that described by Gary T. Marx where individuals take advantage of blind spots inherent in large-scale systems of surveillance. [Marx 2003]

6. IS TRACKMENOT MORALLY DEFENSIBLE

In previous sections we addressed some of TMN's technical limitations, many drawn to our attention by critics. Here, we will discuss challenges to TMN's moral standing. Those we will *not* discuss, however, are accusations that TMN makes life easier for the likes of pedophiles and terrorists by enabling them to hide from view. Although these are important concerns we believe they call attention to the more general challenge of living in a free society where protecting speech, association, and action inevitably creates space for exercising these rights and liberties in ugly and hurtful ways. In order for a society to remain free, it strives to minimize or prevent the hurt and ugliness without diminishing the relevant liberties. This is not a problem for TMN only, nor one that we can make progress on here.

Instead we focus on criticisms addressing specific features of TMN. One such criticism accuses TMN of being no different from 'spamware' or 'Denial of Service' (DoS) attacks, generally wasting network bandwidth and clogging the servers of search engines. Naturally, we resist these critiques. By invoking rhetorical terms like spam and DoS critics seek to cast doubt on our efforts and intentions by associating TMN with activities generally believed to be reprehensible; we see these accusations, however, as question-begging. After consulting numerous sources, we are confident that TMN fits no reasonable, commonly accepted definition of either DoS or spamware. In Wikipedia, for example, a denial of service attack is defined as "an attempt to make a computer resource unavailable to its intended users. Typically the targets are high-profile web servers, and the attack attempts to make the hosted web pages unavailable on the Internet." (Wikipedia, March 29 2007) And spam is defined as "the abuse of electronic messaging

systems to send unsolicited bulk messages, which are generally undesired." (Wikipedia, March 29, 2007) Neither is applicable to TMN.

Behind the rhetoric of DoS and spam, however, there is a question that deserves attention, specifically, the extent of TMN's impact on servers and bandwidth. This concerns us as well, as we had set forth with the value of minimizing resource consumption as a design constraint (see Section 2). The relevant facts are that TMN's resource usage is relatively low, tiny, in fact, compared with common web-traffic such as animations, music and video and consequently it is unlikely to have any appreciable effect on network bandwidth. It is conceivable, however, depending on the number of users and their use patterns (e.g. the mode and frequency settings they select), that TMN might have an impact on search engine performance by placing additional burdens on server processing and bandwidth. Our intention and expectation, based on current usage and trajectory, is that the impact on search engines will be minimal. Universal deployment is not the goal of the project; our intention is to offer a degree of protection to individuals who may feel threatened and afford such users a voice in the evolving debate over web search privacy. We are confident that search companies will take steps to engage with user-dissatisfaction long before TMN usage reaches any significant proportion.

Yet we still have not addressed a key driver of this critique that TMN "wastes" bandwidth and server resources. We know, both anecdotally and through search statistics aggregated by Google Zeitgeist, the Lycos 50 and other such services[12], that people search the Web for a vast range of information. Judging by perennial favorites, the likes of "Britney Spears," "Paris Hilton," and "Pokemon," we conclude that most of these search subjects are not terribly weighty. Further, people and enterprises download and distribute large video, music, and image files of no apparent socially redeeming value and search companies constantly seek out new customers and markets, hoping to entice to their services millions of new users from all over the world. All these activities use bandwidth, place heavier and heavier burdens on servers and services, but they are generally not criticized for wasting network and server resources. Why? These judgments reveal an underlying presumption about what constitutes proper use of the network; these uses, however trivial, are assumed to be legitimate whereas TMN-generated traffic is not. We challenge this assumption. Because adequate privacy protection has been built neither into the technology of search engines, nor the policies governing it, steps taken by

[12] http://www.google.com/press/zeitgeist.html and http://50.lycos.com/

individuals to protect themselves, do constitute a legitimate draw on resources, certainly no less legitimate than the myriad of others drawing on these resources. We place TMN, in this regard, in a category with uses of encryption technologies for securing transactions and proxies for anonymization (e.g. Tor). Despite their incremental draw of resources these additional burdens are generally understood as warranted and at times even necessary. The same goes for TMN.

Another critique charges that TMN is morally indefensible because it violates search engines' Terms of Service (ToS) forbidding access by automated means such as scripts and web crawlers. (e.g. http://www.google.com/accounts/TOS). Although the legal enforceability of website ToS is a broader question than concerns us here, the active debates surrounding it provide valuable input [Burke 1998, O'Rourke 2001, Kerr 2003].[13] Since the very beginnings of the Web, a complex and constantly evolving system of social norms, derived from a combination of law, morality, and affordances of architecture has formed a background against which online actions and transactions are evaluated. Terms of service can be controversial because they unilaterally assert obligations on users that go beyond those implied by the background norms, in particular, some ToS designed to control users' experiences of a website or service. To be sure, website owners' preferences in setting the terms of engagement deserve consideration, but these expressed preferences do not automatically imply moral obligations, particularly ones that society needs honor and defend. Owners' preferences need to be weighed against a range of other considerations.

One such consideration is efficiency. Legal discourse cautions that enforcement of the arbitrary preferences of website owners "for this or that type of usage," [Burk 1998] subjecting users to exclusions and exceptions, would result in the need for users to pick their way cautiously through the Web. Such a requirement would degrade the efficiency and positive externalities of the Web[14]. A Web requiring such cautious engagement is a sadly diminished alternative to the Web extolled for freewheeling access to vast repositories of information goods and services. Beyond efficiency, fairness is another consideration. Since search engines are able to generate value by skimming information

[13] The authors acknowledge the limitations of their perspective on legal issues, formed exclusively with reference to the United States legal system.

[14] Some, like Orin Kerr, have argued that only those ToS expressible in "code" should be enforceable. We imagine protocols like robots.txt would qualify, but the question is a larger, more general one than can be adequately covered here.

off the Web by means of crawlers, benefiting from the willingness of others to place informational resources online with no strings attached, it is unfair to prevent others from doing the same. Fairness precludes making an exception of oneself when one takes full advantage of the norms of open access embraced by others. Legal scholar Dan Burk argues that granting website owners overly strong exclusionary rights would make it possible for them "to free-ride upon the benefits of the network, while at will avoiding contribution of such benefits to others." [Burk 1998]

Although efficiency and fairness are important considerations, they do not necessarily trump the claims of search engines (expressed in ToS) against any and all automated access. Essential to defending TMN is its role in promoting the morally legitimate ends of users' privacy and autonomy in web search and ultimately freedom of expression, association and inquiry. Also relevant is TMN's relatively low imposition on resources. In other words, even if a case could be made to favor the preferences of search engines, as expressed in ToS, against *some* forms of automated query (e.g. ones that are frivolous and seriously undermine performance), fairness and efficiency considerations should place the burden of proof on search companies; in the case of TMN, we believe, they ought not prevail.

7. FUTURE WORK AND CONCLUSION

TMN provides individuals a means of both expressing and asserting a commitment to privacy in web-search without depending on the largesse or intervention of third parties. Although it is fully functional, TMN is best considered a prototype, a proof of concept for a particular approach to privacy, that is, privacy through obfuscation. As discussed, the greatest potential lies in its capacity to protect individuals against profiling and its greatest challenge is to stay abreast of evolving search services themselves. Beyond the challenges of simply keeping up, are challenges of providing rigorous, scientific assessments of performance as well as improving the system in several ways.

A scientific means of evaluating TMN's performance, or the performance of any system adopting this approach needs to address at least one key question, which we are not equipped to answer, namely, how one measures whether user-generated searches have been successfully obfuscated by TMN-generated searches. To be efficacious, TMN needs to introduce into the set of user search queries not only sufficient noise, not only noise in the correct format, but noise of the right kind in relation to the type of protection

being sought and the information being mined. Such needs are likely to turn not only on statistical analysis of signal-to-noise ratios, but also on a practical understanding of how search query data is actually mined and how users are profiled.

Future work to improve TMN could take several directions, mainly focusing on the search query list. One alternative is to incorporate into TMN the means of effectively generating hot button and identifier queries. As mentioned in Section 4, after going some distance along these paths, we chose not to follow them. A second avenue for future work is to find a scalable solution to TMN for languages other than English. Although some among enthusiastic users from non-English speaking countries have offered suggestions and encouragement, a convincing solution is not yet in the offing. A third direction is to explore P2P approaches to generating both search queries and timing patterns as a possible alternative to current mechanisms. A central challenge is to develop a system that meets functional criteria as well as the design constraints discussed in Section 2, such as usability and independence from third parties (i.e. central servers or potentially untrustworthy third-parties.) We are unsure if this is practically achievable.

We conclude with a philosophical point. TrackMeNot operates in an environment that is not only technologically complex, as we have tried briefly to show, but also socially complex. Search engines provide an important service in a volatile and competitive marketplace in which search query logs are a valuable resource and source of revenue. For individuals, however, whether or not they view discrete acts of search and retrieval as sensitive, patterns recorded over time are potentially a window into our lives, interests, and ambitions. As such, they are not only a source of individual vulnerability, but could interfere with free and autonomous inquiry, association, and expression essential to sustaining a healthy democratic society. As a result there remains a tension in the relationship between individual users, important political values, and search service providers. In a better world, this tension would be resolved in a transparent, trust-based mutual accommodation of respective interests.

Instead, users who are concerned with privacy in search perceive little transparency and few credible assurances in the policies of search engine companies that privacy might ever trump pursuit of direct profit. In light of this, trust-based mutual accommodation, of necessity gives way to an adversarial relationship; TMN, a tool for *this* world and this relationship, offers users a say in shaping the terms of engagement with search companies. Although one measure of TrackMeNot's success is impenetrable camouflage

and one hundred percent adoption, we prefer a world in which TMN is no longer needed.

Authors' addresses:
Daniel C. Howe
Media Research Lab
New York University
719 Broadway, 12th floor
New York, NY 10003
dhowe@mrl.nyu.edu

Helen Nissenbaum
Department of Media, Culture and Communication
New York University
239 Greene Street, 7th floor
New York, NY 10003
hfn1@nyu.edu

REFERENCES

---

[1] BARBARO, M. AND ZELLER, T. JR. 2006. A Face Is Exposed for AOL Searcher No. 4417749. In *The New York Times*. 08/09/2006

[2] BIRNHACK, M.D. & ELKIN-KOREN, N. 2003. The Invisible Handshake: The Reemergence of the State in the Digital Environment. *Virginia Journal of Law & Technology* 6:1-57.

[3] BOEHNER, K., DISALVO, C. BODKER, M. AND DE PAULA, R. 2007. The Increasing Value of Reflection: A Discussion of Reflective HCI. Interfaces 72 (2007): 20-23

[4] BURK, D. L. The Trouble with Trespass. 1998. *Journal of Small and Emerging Business Law* 3.

[5] CREATIVE COMMONS. http://creativecommons.org/.

[6] FLANAGAN, M., HOWE, D. C., AND NISSENBAUM, H. 2005. Values at play: design tradeoffs in socially-oriented game design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA, April 02 - 07, 2005). CHI '05. ACM, New York, NY, 751-760.

[7] FRIEDMAN, B., HOWE, D. C., AND FELTEN, E. 2002. Informed Consent in the Mozilla Browser: Implementing Value Sensitive Design. In *Proceedings of the 35th Annual Hawaii international Conference on System Sciences (Hicss'02) - Volume 8* (January 07 - 10, 2002). HICSS. IEEE Computer Society, Washington, DC, 247.

[8] FRIEDMAN, B, KAHN P.H. JR. AND HOWE, D. C. 2000. Trust Online. Communications of the ACM, Vol. 43, No. 12 (2000): 34-40.

[9] FRANCO, R. 2005. Better Website Identification and Extended Validation Certificates in IE7 and Other Browsers. *http://blogs.msdn.com/ie/archive/2005/11/21/495507.aspx, Nov. 21.*

[10] FTC Town Hall Meeting. 2007. Privacy Issues and Behavioral Advertising. November 1-2, 2007, Washington, DC.

[11] HANSELL, S. 2006. Marketers Trace Paths Users Leave on Internet. In *The New York Times*. 09/15/2006

[12] INTRONA, L. AND NISSENBAUM, H. 2000. Shaping the Web: Why the Politics of Search Engines Matters. *Information Society*, 16 (3), 2000, 189-186.

[13] KERR, O. S. 2003. Cybercrime's Scope: Interpreting 'Access' and 'Authorization' in Computer Misuse Statutes. *NYU Law Review 78*, no. 5 (November 2003): 1596-1668.

[14] LATOUR, B. 1992. Where are the Missing Masses? The Sociology of a Few Mundane Artifacts. In *Shaping Technology/Building Society: Studies in Sociotechnical Change*. BIJKER, W.E. & LAW, J. Eds. MIT Press, Cambridge, MA, 225-258.

[15] LESSIG, L. *Code: And Other Laws of Cyberspace*. Basic Books, 1994.

[16] MARX, G. T. A Tack in the Shoe: Neutralizing and Resisting the New Surveillance. *Journal of Social Issue.* Vol 59, No. 2, 2003, 369-390

[17] O'ROURKE, M. A. Is Virtual Trespass an Apt Analogy? *Communications of the ACM* Vol. 44, no. 2 (2001): 98-103

[18] PFAFFENBERGER, B. Technological Dramas. 1992. *Science, Technology, & Human Values*, Vol. 17, No. 3, 282-312

[19] PRIVACY INTERNATIONAL. 2007 A Race to the Bottom: Privacy Ranking of Internet Service Companies,. On *http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-553961* visited Dec. 30, 2007

[20] SCHMIDT, E. 2007. Global Privacy Standards, posted Sept. 19, 2007 On *http://www.peterfleischer.blogspot.com/* - visited Jan. 2, 2008.

[21] TOR: ANONYMITY ONLINE. http://www.torproject.org/

[22] WINNER, L. Do Artifacts Have Politics? 1986. In *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, WINNER, L. Ed. Chicago: University of Chicago Press, 19-39.

[23] Workshop on Ehavioral Advertising: Tracking, Targeting and Technology, November 1-2, 2007

[24] ZIMMER, M. 2007. The Quest for the Perfect Search Engine: Values, Technical Design, and the Flow of Personal Information in Spheres of Mobility. Unpublished Dissertation (NYU).