info@thinkst.com
research@thinkst.com
http://www.thinkst.com

Sock Puppets – Phase One

An Investigation into Building Effective Sock Puppets

This work was sponsored in part by the Open Technology Fund.

Contents

Introduction

About this Document

With the support of the Open Technology Fund and Radio Free Asia, Thinkst Applied Research has investigated the use of online sock puppets in influencing debates and conversations online. The research goal was to explore how fake online personas can be leveraged to suppress or promote agendas, and then to examine the possibilities for timely detection of such activity.

This report details the first phase of our research, that of building sock puppets across a multitude of platforms. In addition to this research report, we also delivered a talk — W*eapons of Mass Distraction: Sock Puppetry for Fun & Profit* — at the Hack in the Box (Kuala Lumpur) conference in October 2014.[1] The slides of the talk are available on the conference website and are included with this paper.[2]

Overview

Three phases are outlined for the research project:

- Phase 1 investigates how to perform measurably effective sock puppetry.
- Phase 2 details empirical evidence and detection of recent (in the wild) sock puppet attacks.
- Phase 3 produces a set of tools for the detection such activity.

This report covers Phase 1. The aim of this phase is to better understand sock puppetry by creating (and measuring) effective sock puppets. Included in this report is a review of notable disclosed examples of sock puppetry and related research.

Research question – Phase 1

How can effective sock puppet campaigns be carried out on mailing lists, Twitter, news websites, online polls and comment systems?

Executive Summary

Phase 1 had two major objectives:

1. Determine if sock puppets could be easily created across major platforms of influence, and
2. Determine if the sock puppets were exerting any influence at all.

Sadly (but somewhat predictably) we were able to achieve the first objective without significant effort. With relatively simple scripting, we were able to create and operate sock puppets on technical mailing lists, popular social networking sites, online polls and the outsourced comment systems that power most of today's news websites.

We largely restricted ourselves to experiments that allowed us to measure the effect of our sock puppetry, allowing us by definition to achieve objective-2, with the ancillary benefit of highlighting paths for improvement and optimization.

From simple mail lists, to social power houses like reddit, from online polls to the comment systems running on CNN, Fox and Al Jazeera one thing remains common: The tools that would allow users, or even system administrators to detect the existence of lame, poorly constructed sock puppets are practically non-existent while a number of reasons ensure that these systems remain laughably easy to reliably influence.

During the course of our work we also discovered security flaws in a number of the services involved. These have been reported to the vendors in question and have either been fixed or are currently in stages of repair.

[1] http://conference.hitb.org/hitbsecconf2014kul/agenda/

[2] http://conference.hitb.org/hitbsecconf2014kul/materials/D2T1%20-%20Haroon%20Meer%20Azhar%20Desai%20and%20Marco%20Slaviero%20-%20Weapons%20of%20Mass%20Distraction.pdf

Background

Context

In the 1920s Walter Lippman highlighted numerous opportunities for the manipulation of the processes which form public opinion, and this is more relevant today considering people spend less effort consciously forming opinions from the information sources available to them.[3] Control over information sources, can and has been leveraged for control over public opinion for many years; of course, the mechanisms of information delivery have changed vastly since the World War I era processes described by Lippman.

Lawyer and legal scholar Tim Wu has identified a common pattern in these changes in the delivery of information, at least in North American information industries such as telephones, radio and film. These industries previously welcomed small dissenting participants, but over time became dominated by monolithic cartels or monopolists. Wu argues that the internet can go a similar way. Dominated by behemoths, they could each have a "master switch" allowing them to kill undesirable content.[4]

This is idea is not farfetched; it is recorded fact that countries have leveraged their positions to disable country-wide internet access when under pressure. In January 2011, Egypt cut internet access almost entirely.[5] The next month internet access in Libya was switched off at the same time as protests on the streets took off.[6] Later that year after protests in England, mobile maker RIM handed over details of devices in the crowd (Blackberries were being used in the organization of demonstrations). Police also used face recognition to identify people in CCTV footage relying on images from Flickr, Tumblr and Twitter.[7,8] These types of abuses are made possible by having vast troves of information access to single (large) entities and are very visible. However, since these examples of censorship are overt and apparent, they are bypassed and subverted once the censorship is discovered.

It is our contention that sock puppets form a new means of shaping narratives and silencing opinions. We define sock puppets as:

> online identities (or personas) created to mislead others by pretending to be different from the operator of the identities, who typically wishes to remain hidden. Sock puppets can vary widely in how realistic they appear, as well as what they are used for.

One example use of sock puppets is to subtly alter information sources without users knowing about the tampering. The proliferation of user generated content (UGC) sites for everything from hotel rankings to Twitter, presents more opportunities for manipulation (and by more people) than those in Wu's master-switch. The intelligent directed use of sock puppets means that dissenting views could be pushed into obscurity, all while the pretense of an open platform for free speech is maintained.

We see this potential dystopia as Censorship 2.0.

Sock Puppetry, A Brief History

This section catalogues historical examples of sock puppetry in order to get a sense for previously seen types of sock puppetry, previous operators of sock puppets and the different observed intentions for the use of sock puppets.

Comment sections of blogs and sites are natural candidates for sock puppetry. A small blogging community was famously (loudly) upset when the author of the *You're Not Helping Blog* was exposed for commenting

[3] Lippman, W. (1922) *Public Opinion* Chapters 2,23,24

[4] Wu, T. (2011) *The Master Switch: The Rise and Fall of Information Empires*

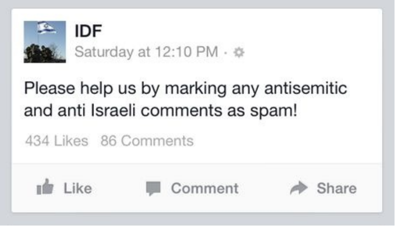
[5] http://www.nytimes.com/2011/01/29/technology/internet/29cutoff.html

[6] http://www.reuters.com/article/2011/02/20/us-libya-protests-internet-idUSTRE71I3XJ20110220

[7] http://www.digitaltrends.com/social-media/london-riots-police-use-flickr-to-help-catch-looters/

[8] http://globalnews.ca/news/143082/the-role-of-digital-and-social-media-in-the-london-riots/

with extra sock puppet accounts to help him win arguments.[9] A more serious case, was that of the user *HamBaconEggs* on the site CommonDreams.[10] HamBaconEggs was one of many sock puppet accounts posting heavily anti-Semetic comments that discouraged some funders from the site. Further investigation exposed the pro-Israel US post-graduate student behind this campaign and brought it to an end. (This highlights that sock puppetry and false flag operations make for common bedfellows.)



IDF
Saturday at 12:10 PM

Please help us by marking any antisemitic and anti Israeli comments as spam!

434 Likes 86 Comments

A recent post from an Israeli Defense Force (IDF) Facebook page

Apart from polemics, sock puppets have also been used for commercial gain. In 2004 a bug in the Amazon online store leaked private information about users that leave book reviews.[11] This exposed both the writers who had written glowing reviews of their own books, and those who had left disparaging reviews on competing titles. Another well documented example was with the link-aggregation site Reddit. When it first began, the site administrators made use of sock puppet accounts to post content creating the impression that the site was more active than it actually was.[12] With similar aims of bolstering an organization, the Fox News PR department ran a sock puppet campaign over a number of years.[13] A former employee leaked details of having operated at least 100 sock puppet accounts to counter blogs posts and comments that were critical of the network.

Governments and political parties have also employed similar tactics to the Fox News PR department. The Chinese government is said to run a team of paid commenters called the 50-cent army.[14] The Haaretz documented the use of teams of Israeli students, incentivized by the office of the Prime Minister to comment positively about Israel online.[15] There were also signs of Russian online commenting teams[16] as well as one run by the Turkish ruling party, AKP.[17]

There are a few more interesting cases of state-sponsored sock puppetry that are somewhat different to simple online commenting teams. In 2010 the United States Central Military Command (CENTCOM) solicited bids for “Persona Management” software to do sock puppetry that stringently masked the operator’s identity.[18] In another interesting case, Rwanda retained the PR firm Racepoint in 2009 to promote the entire country and encourage investment in country. The agreement signed between the two included publishing articles with media houses, and continuously posting and promoting these on content aggregations sites such as Digg and Reddit.[19]

Intelligence services have also made use of sock puppetry. In the lead up to the 2012 South Korean elections, the national intelligence service of South Korea, ran a Twitter campaign to influence the outcome the elections. At least 1.2 million tweets went out during the campaign in an attempt to smear the major opposition party.[20] On the other side of the world, the Snowden docs revealed that GCHQ’s JTRIG group had a programme dedicated to gaming online polls (called “UNDERPASS”).[21]

Another hybrid option for sock puppetry is to encourage real users to do a something on a platform. This is instead of creating sock puppet accounts that are managed by an operator. Wikipedians call it “Meat puppetry”[22] when a band of users respond positively to encouragement to make sure an article is edited a particular way. Recently the Israeli Defense Force put out a similar type of appeal on Facebook, encouraging users to mark anti-Israeli comments as spam to get them removed.

[9] http://thebuddhaisnotserious.wordpress.com/2010/06/19/the-curious-case-of-the-youre-not-helping-blog/
A poem was also written about the incident: http://quichemoraine.com/2010/06/the-saga-of-the-youre-not-helping-blog/

[10] http://www.commondreams.org/hambaconeggs

[11] http://www.nytimes.com/2004/02/14/us/amazon-glitch-unmasks-war-of-reviewers.html

[12] http://venturebeat.com/2012/06/22/reddit-fake-users/

[13] http://mediamatters.org/blog/2013/10/20/fox-news-reportedly-used-fake-commenter-account/196509

[14] http://www.newstatesman.com/politics/politics/2012/10/china%E2%80%99s-paid-trolls-meet-50-cent-party

[15] http://www.haaretz.com/news/national/.premium-1.541142

[16] http://www.theatlantic.com/international/archive/2013/10/russias-online-comment-propaganda-army/280432/

[17] http://online.wsj.com/news/articles/SB10001424127887323527004579079151479634742

[18] http://echelon2.org/wiki/Persona_Management

[19] http://www.fara.gov/docs/6055-Exhibit-AB-20110812-1.pdf

[20] http://www.nytimes.com/2013/11/22/world/asia/prosecutors-detail-bid-to-sway-south-korean-election.html?_r=1&

[21] https://firstlook.org/theintercept/2014/02/24/jtrig-manipulation/

[22] https://en.wikipedia.org/wiki/Wikipedia:Meat_puppetry

A final example, shows a very different use of sock puppets. Jeff Bardin describes in a talk[23] how he creates and manages sock puppets to infiltrate suspected terrorist groups. These sock puppets are exquisitely crafted with rich individual histories and have typically develop long lasting relationships with targets (and with each other).

Clearly sock puppetry is used by different groups from lone students to intelligence services and for different reasons.

Related Work

Aside from isolated incidents documenting particular sock puppets, there has been some research on the general question of detecting sock puppets (and related problems). Detecting shills or astroturfing—(where the appearance of grass-roots support is faked)—and determining the credibility of user generated information have also been studied. One of the difficulties with proposing a method for the detection of any of these, is measuring the effectiveness of the detection. The more convincing research to date, tends to rely on manual verification of the results (vs. detection methods).

Researchers have highlighted the need to detect sock puppetry online in different ways. Two US researchers argued that obscure views can be made disproportionately prominent in a matter of minutes when they examined a Twitter political campaign in 2010. They note that lots of fabricated content (such as those contributed by sock puppets) are picked up and promote by search engines.[24] Another researcher points out that the "automatic detection of sock puppets" is a useful and necessary line of research in order to get a reliable interpretation of tweets.[25] The point can be made more generally that user-generated content on an online platform should be used with caution whether by verification or by detecting sock puppet activity. [26]

One of the key difficulties in tackling the automatic detection of sock puppets, is being able to verify that the output of the results are actually sock puppets. Some attempts propose algorithms and then manually verify that the identified accounts are in fact sock puppets,[27] but this is limited to the number that can be manually verified. Another difficulty is determining the efficacy of the algorithm without knowing what the rate of false negatives is.

To get around the limited number of manual verifications the Truthy Project,[28] which aims to detect astroturfing, uses crowd-sourcing to identify non-organically created memes on Twitter. This data is also used to further train their machine learning algorithm after the initial training data. Tweetcred[29] takes a similar approach to detecting the credibility of tweets. Both projects manually identify many variables, and leave the learning algorithm to determine their importance in the judgment.

A slightly different approach aims at providing people with tools to make it easier for people to detect sock puppetry, instead of a completely automating the detection. WikiWatchdog[30] helps with this on Wikipedia by showing article edits made from IP addresses, to see the range that they cover. WikiWatchdog can help notice abuse beyond simple, shallow sock puppets. Some commenting systems also provide basic tools like this to help admins detect suspicious behavior and simple sock puppetry.[31] The HamBaconEggs case mentioned in the previous section used the tools available to Disqus comment moderators to track down the sock puppet operator.

Research into the automatic detection of sock puppetry, has had some success in finding sock puppet accounts, though there is not a reliable method of completely assessing the efficacy of such research (particularly in determining the false-negative rate of detection methods). The slightly different approach of providing tools to help users detect sock puppetry has been less explored, but has already shown real world benefits.

[23] http://privacy-pc.com/articles/open-source-intelligence-by-jeff-bardin.html

[24] http://journal.webscience.org/317/2/websci10_submission_89.pdf

[25] http://arxiv.org/pdf/1204.6441v1.pdf

[26] http://irevolution.files.wordpress.com/2011/11/meier-verifying-crowdsourced-data-case-studies.pdf

[27] http://www.cs.hku.hk/research/techreps/document/TR-2011-03.pdf

[28] http://www.truthy.indiana.edu/about

[29] http://arxiv.org/abs/1405.5490

[30] http://www.wikiwatchdog.com/

[31] https://help.disqus.com/customer/portal/articles/466238-moderating-your-community

Experiments and Results

General Approach

It is worth reiterating the aim of Phase One: based on our hypothesis that sock puppets will be used as a means of hijacking online conversation streams and drowning out competing voices as a form of censorship 2.0, we look at how attackers (private or public) can use sock puppets to influence discussion or other sources of information. We examine ways of silencing activists and dissenting views, or promoting a desired view all while maintaining a pretense of supporting free speech on internet platforms.

For this phase we built sock puppet software and used it to influence conversations happening on popular sites. The two key challenges here are:

1. building the sock puppets (and overcoming any defenses on each platform), and
2. determining whether the sock puppets are having any *measurable* meaningful effect.

Demonstrating that the sock puppets actually had an effect on the targeted users proved to be a bigger challenge than expected. We settled on a series of experiments to test individual hypotheses, in order to allow for metrics gathering and to enable repetition of results. Data gathered in these experiments allow us to make statements and draw conclusions with more certainty on the sock puppetry's effectiveness, which is in contrast to the often vague and untested claims that tend to accompany discussions on this topic.

The experiments fell into two general groups:

1. Those that influenced attention.
2. Those that influenced location.

In the first case, link clicks were commonly used as a metric to infer attention from. Experiments in group 1 measured whether sock puppet activity was able to influence link clicks (i.e. whether more or less people have seen a candidate link). Both attracting and diverting attention are important.

In the second case, we aim to determine whether we can reliably influence what shows up in visual spaces of prime importance on specific platforms, such as "Most read" lists.

We chose categories of sources we believed would be worth examining: news sites, user link submission sites and mailing lists, and from these categories selected specific channels We list them below:

Category	Samples
Mailing Lists	LiberationTech FullDisclosure
Online Polls	Polldaddy Huffington Post
Blogging	Twitter
Link aggregators	Twitter HackerNews
News	Mail & Guardian Wall Street Journal The New York Times
Comment hosting	Disqus Livefyre

The rest of this section covers the experiments and their results.

Mailing Lists

Internet mailing lists have a long and illustrious history, dating back to at least 1986.[32] They emerged as the central coordination mechanism in the development of fundamental Internet standards, software and projects, in addition to the huge variety of private and public lists on virtually every topic imaginable. Open source software development in particular is heavily reliant on mailing lists, and Internet standards discussions still take place on mailing lists.

Mailing lists are attractive for sock puppetry for three reasons. Firstly, email addresses are essentially free and personas are easy to create, meaning that puppets can be created at will. Secondly, sending email is also essentially free, so a puppet master expends very few resources in transmitting a message. Lastly, important discussions still occur via mailing lists.

Examples of current important discussions are the standardization of encryption ciphers, the adoption of new random number generators, the inclusion of DRM technology in web browsers, and security design choices in operating systems. A puppet army could inflate the perceived support of some design choice, or inflate its opposition. A puppet army could also drive open source developers to distraction causing them to abandon paths or software, especially if the development is after hours.[33]

Aims

The goal of mailing list sock puppetry is to either promote or suppress an email, depending on the views of the handler. Specifically this means using sock puppet to either attract or divert attention by getting:

- more people to see an email, or
- less people to see an email.

As described earlier, attention is measured via link clicks so each email included a link that we controlled and every hit to the link was recorded. This serves as a proxy for the amount of attention an email gets.

Approach

Each of the two aims was tested by a separate experiment.

Experiment 1: Attract attention with discussion thread on a target email

We propose that attention can be attracted to an email, by creating a faked discussion that responds to the target email. Most email clients will show responses (and responses to responses) in a tree or threaded view. This tests whether a target email will get more views, if sock puppets create a long discussion centered around the target email.

We compare the attention received by an email sent without a fake thread (the control email) versus an email sent along with a sock puppet-fueled discussion thread.

The experiment runs as follows:

Control	Experiment
1. Send the control email with a link 2. Count the number of link clicks	1. Send the experimental email with a link 2. Send several sock puppet emails to generate a discussion thread on that email 3. Count the number of clicks.

After each email with a link, we waited 48 hours for the link clicks to die down, before counting the total.

[32] LISTSERV Archive, http://peach.ease.lsoft.com/scripts/wa-peach.exe?A1=ind8607&L=LSTSRV-L

[33] A prime target currently is the author of the grsecurity Linux Kernel patch, who maintains the most comprehensive set of security strengthening patches in his own time. If he grows tired of the project, the Linux ecosystem will be much poorer for it.

Experiment 2: Divert attention with competing threads

This tests whether a candidate email receives less views if our sock puppets start several distracting or competing email threads directly after the candidate email arrives. Again the comparison is between an email sent alone (control email) versus the candidate email sent, and followed by several emails starting new discussion threads.

The experiment is run similarly as the first:

Control	Experiment
1. Send the control email with a link 2. Count the number of link clicks	1. Send the experimental email with a link 2. Sock puppets start distracting threads after target email. 3. Count the number of clicks.

Experiment venues

We ran the two experiments on two public discussion mailing lists with relatively technically sophisticated users:

- LiberationTech,[34] a list on the use of technology in society for public benefit, and
- FullDisclosure,[35] a technical security list on vulnerabilities, exploits and general topics in the computer security community.

LiberationTech moderates signups despite performing no verification of people joining the list. (In fact they encourage anonymous signups.) However, once joined, a user is able to post without any moderation.

FullDisclosure by contrast allows instantaneous signups, but every email sent to the list is moderated. This produces an unpredictable delay in the distribution of a sent email (and the moderators sometimes reject emails which adds to the uncertainty).

We used link clicks in plaintext emails to allow comparing the results between the two mailing list, as FullDisclosure strips HTML emails, and will be seen by more users (some users may only view the plaintext portion.) The content in the email was written by hand and we focused on making it enticing as well as appropriate for each setting. Generally the topic of the email is copied from elsewhere.

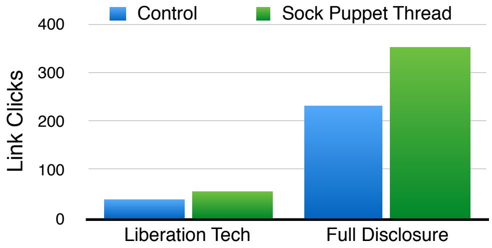
The email addresses used were a combination of trial accounts from an email hosting service and disposable addresses.[36] (Emails were spoofed from the disposable email addresses.)

Results

Both experiments succeeded in showing a difference between the control email and the experimental email, in both attracting and diverting attention. One confounding variable was that occasionally the control email was sufficiently interesting to generate long discussion threads, but this was not often observed. LiberationTech proved to be the easier test venue as emails did not require moderator approval.

Experiment 1: Attracting attention

On both mailing lists the email with the fake discussion thread received more clicks than the control mail as shown in the chart on the right which shows the results of a run of the experiment.



Link clicks for a run of the fake discussion thread experiment

Liberation Tech is a lower profile list and has fewer subscribers then Full Disclosure, which is reflected in the difference in absolute counts for the control and sock puppet clicks. Regardless, the attention increase was observed across both mailing lists, when a discussion thread was started on the target mail.

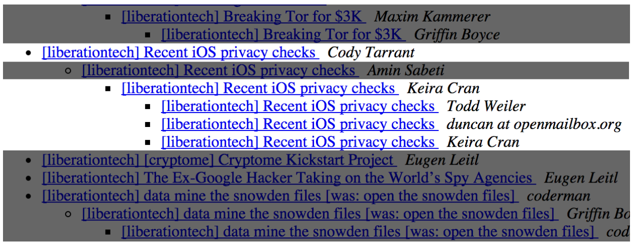
The discussion threads do not need to be long in order to

[34] https://mailman.stanford.edu/mailman/listinfo/liberationtech

[35] http://nmap.org/mailman/listinfo/fulldisclosure

[36] http://www.mailinator.com/

attract attention; the screenshot below shows a successful fake thread with only a handful of sock puppet replies:



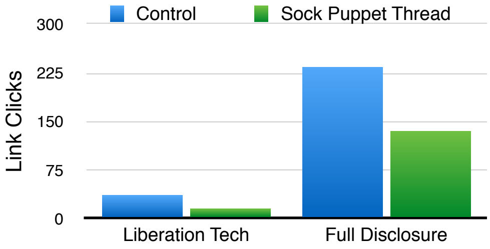
        - [liberationtech] Breaking Tor for $3K *Maxim Kammerer*
            - [liberationtech] Breaking Tor for $3K *Griffin Boyce*
- [liberationtech] Recent iOS privacy checks *Cody Tarrant*
    - [liberationtech] Recent iOS privacy checks *Amin Sabeti*
        - [liberationtech] Recent iOS privacy checks *Keira Cran*
            - [liberationtech] Recent iOS privacy checks *Todd Weiler*
            - [liberationtech] Recent iOS privacy checks *duncan at openmailbox.org*
            - [liberationtech] Recent iOS privacy checks *Keira Cran*
- [liberationtech] [cryptome] Cryptome Kickstart Project *Eugen Leitl*
- [liberationtech] The Ex-Google Hacker Taking on the World's Spy Agencies *Eugen Leitl*
- [liberationtech] data mine the snowden files [was: open the snowden files] *coderman*
    - [liberationtech] data mine the snowden files [was: open the snowden files] *Griffin Bo*
        - [liberationtech] data mine the snowden files [was: open the snowden files] *cod*

Original email thread with non-puppet mails greyed out. The first puppet mail is the target, the remainder add weight.

Experiment 2: Diverting attention

For the distraction threads experiment we observed that emails followed by multiple distractor threads received less clicks than their control counterparts. Again the chart to the right shows the results of a run of this experiment.

Apart from the difference in absolute numbers (explained above), both mailing lists exhibit a similar drop-off in attention when the sock puppetry technique is applied, compared to the control. The experiment was successful, and repeated multiple times to confirm the results.



Link clicks for a run of the distraction threads

Discussion

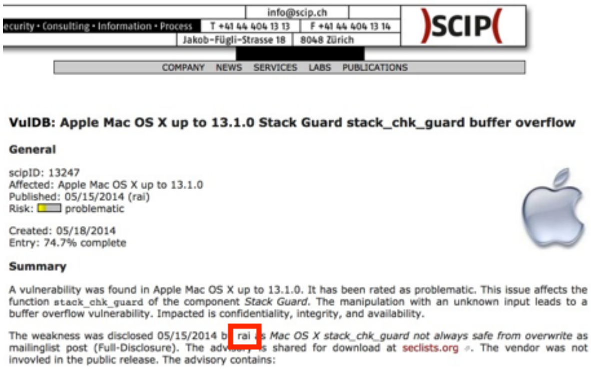
These results show that both attracting and diverting attention are possible on mailing lists using the mechanisms of thread lengthening and thread multiplication. Writing the distraction threads ahead of time on topics known to excite the mailing list members can increase the efficacy the distraction.

Our more successfully distraction runs had significant organic participation in the distraction threads. This suggests an obvious extension: suppress emails by immediately having sock puppets follow them up with distraction emails. Each of those distraction emails, should have more sock puppets creating fake discussion threads, to attract people to the distraction discussion threads. These threads can be prepared ahead of time, to be on a topic of interest to the list.

Despite how simple mailing lists are, there still seems to be room for influencing them with sock puppet email accounts. There were no signs of anyone noticing that different email addresses domains were sent from the same IP for emails that were sent close together, nor that disposable email addresses were used as the From address for some mails. Even then, getting more email addresses is straightforward.

Another way of drawing attention to an email, could be to start very obvious sock puppet distraction threads after an email we wish to promote. The aim would be to be have the sock puppets be easily discovered, and leverage the Streisand effect[37] when people notice that the email that was "suppressed" and in doing so, promote the original email.

Interestingly, during one (failed) run of the experiment on Full Disclosure, the "author" of our control email (named after a cat) ended up being cited as a security researcher in a small vulnerability database on the Internet.[38] This idea of stealing good content from elsewhere to boost your own image crops up again later and is a useful technique for adding "personality" to fake personas.



ecurity • Consulting • Information • Process | T +41 44 404 13 13 | F +41 44 404 13 14 | info@scip.ch | Jakob-Fügli-Strasse 18 | 8048 Zürich

)SCIP(

COMPANY NEWS SERVICES LABS PUBLICATIONS

VulDB: Apple Mac OS X up to 13.1.0 Stack Guard stack_chk_guard buffer overflow

General

scipID: 13247
Affected: Apple Mac OS X up to 13.1.0
Published: 05/15/2014 (rai)
Risk: ▭ problematic

Created: 05/18/2014
Entry: 74.7% complete

Summary

A vulnerability was found in Apple Mac OS X up to 13.1.0. It has been rated as problematic. This issue affects the function `stack_chk_guard` of the component *Stack Guard*. The manipulation with an unknown input leads to a buffer overflow vulnerability. Impacted is confidentiality, integrity, and availability.

The weakness was disclosed 05/15/2014 by rai as *Mac OS X stack_chk_guard not always safe from overwrite* as mailinglist post (Full-Disclosure). The advisory is shared for download at seclists.org. The vendor was not invovled in the public release. The advisory contains:

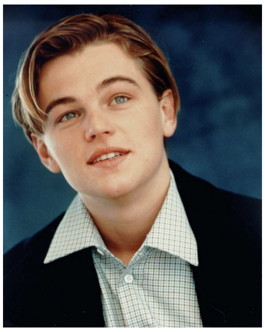
A screenshot of the listing in the vulnerability database citing our sock puppet email author "rai."

[37] http://www.economist.com/blogs/economist-explains/2013/04/economist-explains-what-streisand-effect

[38] http://www.scip.ch/en/?vuldb.13247

Online Polls

Polls (of the click-to-vote variety) have been around for a while. Typically these are opinion polls on news websites, small polls on blogs, and are occasionally employed as part of a larger process (such as selecting the TIME magazine person of the year). Poll gaming has been around for a long time and one early example is the People magazine's "Most Beautiful Person" poll of 1998: Leonardo di Caprio at the height of his Titanic adulation garnered an impressive 14 thousand votes but lost out to entertainer Hank, "the Angry Drunken Dwarf" with 230 thousand votes.[39]



Two contenders for the 1998 People magazine "Most Beautiful Person" poll: Hank the angry drunk dwarf (left) and Leonardo di Caprio (right).

Aims

Our goal was to manipulate widely seen polls and poll services. In particular, make a desired poll option:

- win subtly by narrowly beating the next option, or
- win by a huge landslide.

The landslide win, if done clumsily, could be useful as a false flag. The attraction of polls is that their results are used, particularly if viewed on high-traffic news sites.

Approach

We examined PollDaddy, which was at one point used by TIME for their person of the year poll, and many blogs worldwide. For widely seen polls, we examined the Huffington Post Readers' Poll and a poll on the front page of Al Jazeera's Arabic news site.

Results

Polldaddy

Given how widely used PollDaddy is, there are many tutorials, videos, and tools available for gaming the polls in their default configuration. Most of the material is outdated and no longer works; one can infer from this that gaming these polls has been going on for a while. A blog post describing how the latest iteration could be fooled worked during our research, making verification trivial.[40] A poll manipulator need only fetch a nonce from the server before voting, meaning two requests are needed per vote. As a result we were able to script the landslide win within minutes, and modified it slightly to also be capable of more subtle wins.

If different IP addresses were required for votes (as they sometimes may be) we used open web proxies. There are other poll configurations that require an email verification or account creation, however, these are hardly ever seen and are not available for the vast majority of non-paying PollDaddy users. The interface for paid users, does include rudimentary interface to see how many votes came in from which IP addresses and affords poll admins the ability to discard those that look suspicious.

Huffington Post

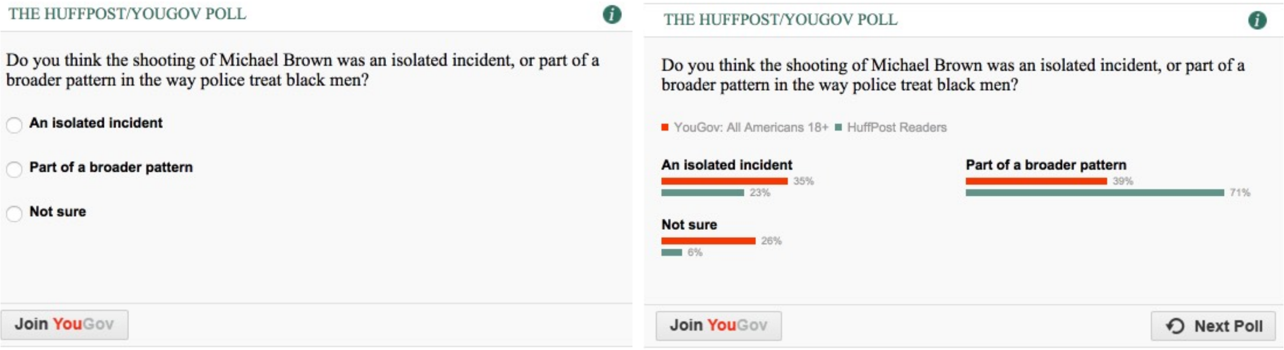
YouGov is a market-research firm that relies on Internet surveys to conduct their research. Participants are users who have previously signed up to the YouGov site and filled in biographical forms. With this database of users and their preferences and history, organizations approach YouGov to target questionnaires are whatever demographic is of interest. For example, a gaming firm can easily survey single males with a gaming interest between the ages of 15–25 via YouGov.

[39] https://en.wikipedia.org/wiki/Hank_the_Angry_Drunken_Dwarf

[40] http://codeantics.wordpress.com/2013/12/12/poll-daddy-reverse-engineering/

Participants are rewarded with points which can eventually be cashed out. The selected participant model is claimed to be more accurate at predicting polls than the traditional polling methods.

News website The Huffington Post often embeds a readers' poll on topical news stories alongside results from a YouGov survey on the same issue. After voting, the results of the poll are shown to the user. In the figures below, the green bar shows the results of the Huffington Post poll and the red is the separate survey run by YouGov on their members.



THE HUFFPOST/YOUGOV POLL

Do you think the shooting of Michael Brown was an isolated incident, or part of a broader pattern in the way police treat black men?

○ An isolated incident

○ Part of a broader pattern

○ Not sure

Join YouGov

THE HUFFPOST/YOUGOV POLL

Do you think the shooting of Michael Brown was an isolated incident, or part of a broader pattern in the way police treat black men?

🟥 YouGov: All Americans 18+ 🟩 HuffPost Readers

An isolated incident 35% 23%

Part of a broader pattern 39% 71%

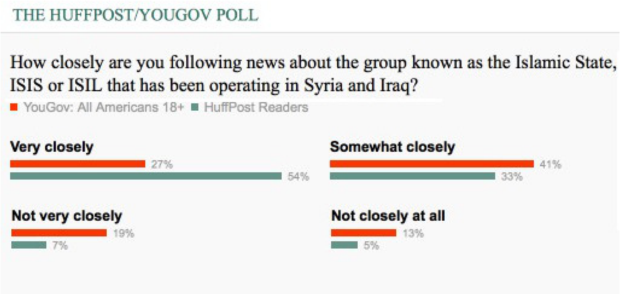
Not sure 26% 6%

Join YouGov

Next Poll

Screenshots of voting in the Huffington Post Readers' Poll and viewing the results

This readers poll is trivial to game. We were able to vote repeatedly from a single IP address with a one line script. As a practical aside, we voted from Amazon EC2 machines, and needed four instances running in parallel to get enough horsepower to return a landslide win. The images below show the poll before and after we voted on the option: "Not Closely At All". The green bar shows the poll we participated in, as the percentages grow from 5% to 65%. For a subtle win, the script just monitors the previous poll results and votes on it periodically.



THE HUFFPOST/YOUGOV POLL

How closely are you following news about the group known as the Islamic State, ISIS or ISIL that has been operating in Syria and Iraq?

🟥 YouGov: All Americans 18+ 🟩 HuffPost Readers

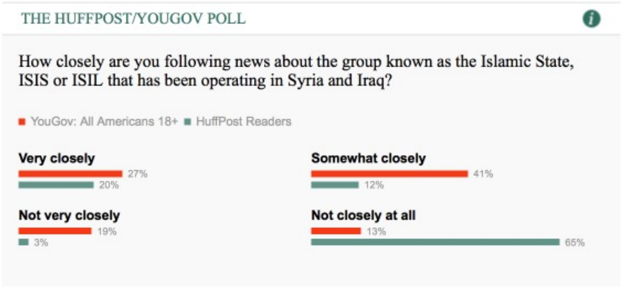
Very closely 27% 54%

Somewhat closely 41% 33%

Not very closely 19% 7%

Not closely at all 13% 5%

The results of a Huffington Post poll before gaming it. The option "Not closely at all" has 5% of the votes.



THE HUFFPOST/YOUGOV POLL

How closely are you following news about the group known as the Islamic State, ISIS or ISIL that has been operating in Syria and Iraq?

🟥 YouGov: All Americans 18+ 🟩 HuffPost Readers

Very closely 27% 20%

Somewhat closely 41% 12%

Not very closely 19% 3%

Not closely at all 13% 65%

The results of the Huffington Post poll after gaming it. The option "Not closely at all" now has 65% of the votes.

Discussion

More than a decade after the gaming after the "Most Beautiful Person" poll, we found it surprising that the Huffington Post's readers poll (and indeed most of the others examined) were so trivial to game. There are viable ways of improving poll reliability while still keeping the voting easy enough for users but this does not appear to be in common usage at all. Simple anti-automation steps work wonders for defeating trivial gaming attempts.[41] However, the typical poll on the internet, particularly a PollDaddy one, shouldn't be relied on to be representative of the views of users.

Twitter

Twitter's popularity means it needs no introduction. This was underlined in 2013 when stock markets dipped in response to tweets from the Associated Press which claimed that there were explosions in the WhiteHouse; the account had been take over by attackers and was issuing false information.

Twitter is an attractive target for sock puppetry due to the way it is consumed. Instead of inbox style reading (like mail or RSS) where the intent is to read every item, Twitter is intended to be consumed by "dipping into the stream" or reading the most recent tweets on your stream. This allows a user (or a number of users) draw attention to a topic by being extra verbose or by sending out a number of tweets in rapid succession

[41] An example is Al Jazeera (Arabic) which regularly receives a few thousand votes on its polls despite users having to enter a CAPTCHA before voting.

when she wishes to distract you. These steps are simple, but premised on an important fact: the target needs to follow the sock puppet.

The most important step in Twitter sock puppetry is therefore by convincing the target to follow the puppet, and our experiments primarily dealt with this aspect.

Aims

The primary investigation into Twitter examined strategies for getting follow backs. To achieve this we needed to:

- investigate the purchasing of twitter followers;
- use bought followers to obtain real followers we wish to influence.

Approach

Gilad Lotan documented his experience with buying followers and demonstrated that even though the quality of bought followers was terrible, a higher follower count appeared to lead to a disproportionate growth in organic followers. In other words, bought followers attracted real followers.

We experimented with this by purchasing a few thousand followers but were unable to successfully replicate these results.

Results

All the experiments we chose to run on Twitter required weeks of setup and execution, and in the final analysis, we fared poorly in scoring any useful victories here. We do not consider the results evidence of impossibility, and are cautiously optimistic that effective Twitter sock puppetry is likely.

Reddit

Reddit began as a small user-driven content submission and aggregation site. Ironically, in its early stages Reddit was so small that the site administrators would post extra content as fake users[42] but by June 2014 the site saw over 8 million unique visitors.[43] Today there are multiple independent services dedicated to influencing the site (500 votes can easily cost $300–$450) and Reddit features on the selling page of many “social media experts”.

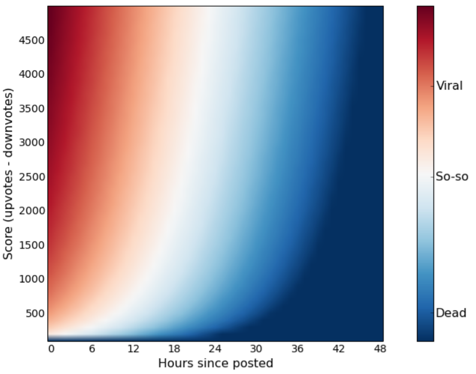
Redditors submit, up-vote, down-vote, and comment on posts and links to various subreddits on the site. Subreddits are separate areas within the site dedicated to a huge range of topics, and they are managed by regular users called moderators, rather than Reddit employees. The front page of a subreddit features a list of the current most popular posts. (There are other ranking methods but these are not displayed by default.) Featuring on a widely read subreddit front page guarantees views on a post.

Aims

On Reddit our aims are to influence a subreddit front page using sock puppet accounts. Specifically these are:

- to promote posts onto the subreddit front page; and
- to suppress posts and keep them off the subreddit front page.

The default ranking is by the post’s hot score which is a function of the age of the post (time) and the vote score (up-votes less down-votes). Newer posts feature higher than older posts with the same vote score. Older posts require exponentially more votes (relative to the time difference) to compare favorably with newer posts. The graph to the right depicts the drop off in the hot score (red-blue) as a post ages, despite having lots of votes.



Reddit Hotness by Score and Age

[42] http://www.dailydot.com/business/steve-huffman-built-reddit-fake-accounts/

[43] https://siteanalytics.compete.com/reddit.com/

[44] http://www.randalolson.com/2014/03/21/the-window-of-virality-on-reddit/

So to influence the default hot score page, the aim here is to game the score of posts through the two important factors: age of the post and the vote score. Since the age of the post is out of our control for a post we have not submitted, the focus is on using voting to influence the ranking of the articles.

Approach

Our Reddit experiments were multi-faceted and consisted of a number of sub-steps. In order to sock puppet, we needed accounts and voting strategies. In this line, we needed to:

- create *effective* sock puppets,
- get a post onto the subreddit front page by up-voting,
- get a post off the subreddit front page by down-voting, and
- control what features on a subreddit page.

From the range of subreddits to chose from, most of our experiments were run on the */r/worldnews,*[45] and */r/netsec* subreddits.[46] */r/worldnews* has a popular audience of 6 million subscribers for its content of links to non-US news stories and */r/netsec* has a smaller audience of around 100k subscribers, but who tend be technical-minded.

Results

Creating Reddit users

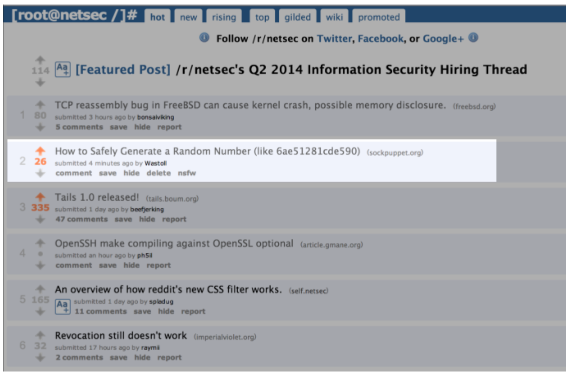
The first challenge was to create a collection of sock puppets. There were two initial hindrances to automating account registration:

- registrations from a single IP were rate-limited to 1 registration every 10 minutes, and
- successful registration required solving a CAPTCHA.

To bypass the IP rate limiting, we once again relied on open web proxies to register accounts from different IP addresses. These proxies are easily found on publicly available lists on several sites.[47] The CAPTCHA were solved manually.[48]

Up-voting

Reddit has a long history of users voting with sock puppet accounts and the site does throw up a few defenses. The most notable is that Reddit obscures the true vote score of a post by showing a "fuzzed" vote score that is apparently not the actual vote score. In other words, if a post has 10 up-votes and 6 down-votes, its score is 10–6=4, but Reddit will show a score anywhere between 2 and 8 on subsequent refreshes. The intention is so that a sock puppet cannot immediately tell if their vote counted. Additionally, when the platform detects suspicious up or down-votes, an extra vote is automatically placed in the opposite direction. This keeps the vote score the same, even though the vote has been counted. Lastly, a user can be "shadow-banned" where the effects of their actions are only visible to themselves, but not to other users of the sites (effectively nullifying any of their votes).



Follow /r/netsec on Twitter, Facebook, or Google+

114 [Featured Post] /r/netsec's Q2 2014 Information Security Hiring Thread

1 80 TCP reassembly bug in FreeBSD can cause kernel crash, possible memory disclosure. (freebsd.org)
submitted 3 hours ago by bonsaiviking
5 comments save hide report

2 26 How to Safely Generate a Random Number (like 6ae51281cde590) (sockpuppet.org)
submitted 4 minutes ago by Wastoll
comment save hide delete nsfw

3 335 Tails 1.0 released! (tails.boum.org)
submitted 1 day ago by beefjerking
47 comments save hide report

4 • OpenSSH make compiling against OpenSSL optional (article.gmane.org)
submitted an hour ago by ph5il
comment save hide report

5 165 An overview of how reddit's new CSS filter works. (self.netsec)
submitted 1 day ago by spladug
11 comments save hide report

6 32 Revocation still doesn't work (imperialviolet.org)
submitted 17 hours ago by raymii
2 comments save hide report

Screenshot of the netsec subreddit front page highlighting a post we voted into second place.

The result of these defenses is that after a casting a vote from one user, it is not immediately obvious whether this single vote has had an effect by looking at the scores. Initial testing with several accounts failed to narrow down an exact set of account characteristics (e.g. email verification, age of accounts) that would allow an account's votes to count. However, as it turned out this was not necessary for the experiments we wanted to run.

With a group of 50 accounts, we ran the first "up-voting a

[45] http://www.reddit.com/r/worldnews

[46] http://www.reddit.com/r/netsec

[47] See for example http://www.xroxy.com/proxylist.htm

[48] CAPTCHA solving could be farmed out to services but even manually we were able to create a few hundred accounts in a day.

post” experiment to determine whether we had any impact on the actual ranking of a post. (This is in contrast to trying to determine the effect on the actual hot score.) The experiment to up-vote a post on the front page was successful on the */r/netsec* subreddit. By posting a new post (and employing the help of just a small number of our bots) we were able to consistently up-vote our article onto the front page. On */r/worldnews*, despite having some visible effect, our influence was not large enough to promote an article onto the front page but moved articles lower down the rankings (e.g. pushing articles from 70 to 30). This makes sense, as the number of other users voting on posts in */r//worldnews* is much larger. This however is simply a matter of scale. By increasing the number of bots (and machines running our bots), we have no reason to believe that controlling the front page */r/worldnews* would be any different from */r/netsec.*

Down-voting

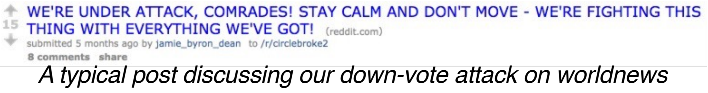
Next we experimented with down-voting a single post. Depending on the subreddit configuration two interesting possibilities can happen after a few (not tagged as spam) down-votes:

1. the post gets removed from the subreddit, and added to the moderation queue for approval before being displayed again,
2. the post becomes invisible to many users, because of a default user-defined threshold for which posts below a certain score are no longer visible. (The default is –4.)

Abusing the post-visibility threshold can be a viable tactic for suppressing newly created posts. Without earning any up-votes early on, the post will languish in obscurity as it ages.

Mass down-voting

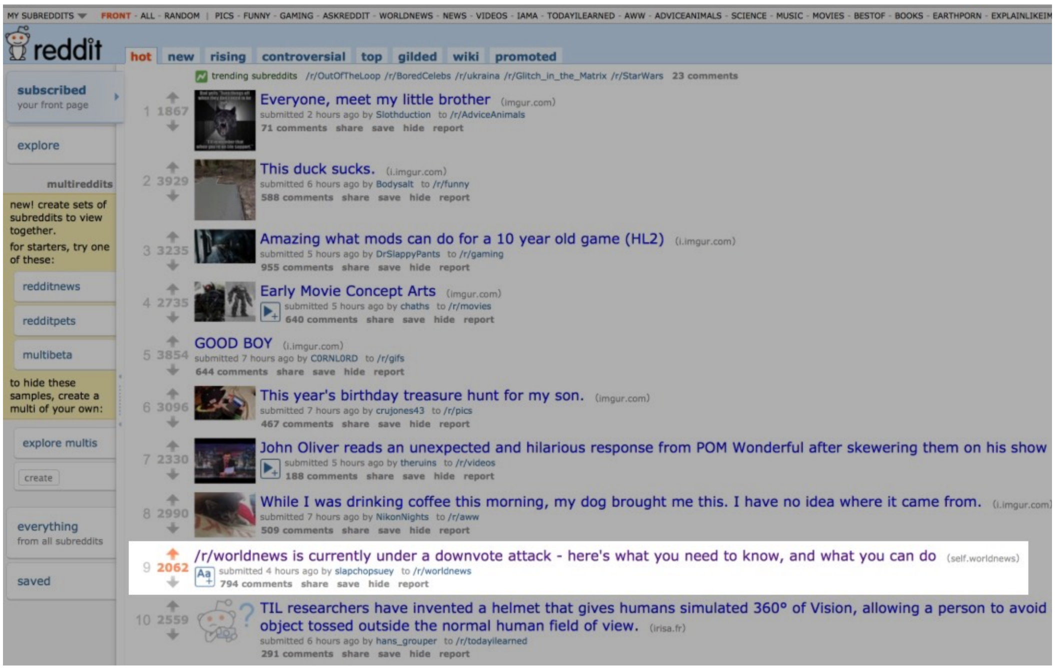
The next experiment we tried was to keep all new posts off the front page of the subreddit by simply down voting all new articles on both */r/worldnews* and */r/netsec*. The result was that the mass down-voting did not keep the posts off the front page.



15 WE'RE UNDER ATTACK, COMRADES! STAY CALM AND DON'T MOVE - WE'RE FIGHTING THIS THING WITH EVERYTHING WE'VE GOT! (reddit.com)
submitted 5 months ago by jamie_byron_dean to /r/circlebroke2
8 comments share

A typical post discussing our down-vote attack on worldnews

However the effect of this simple experiment was interesting: almost immediately users on */r/worldnews* began to notice that each new post had a huge number of down-votes, and this sparked discussions over the down-voting. Several hours later, the problem came to the attention of the */r/worldnews* moderators who posted a notice about the attack and mentioned that the that the administrators of Reddit had been notified. The unstated implication was that the moderators could do nothing, and were going to wait it out. This post itself became very popular, making 9th position on the Reddit’s front page with around 1700 comments. (This page shows the most popular posts across all the subreddits on Reddit. Very few posts make it onto this page.)


MY SUBREDDITS
FRONT - ALL - RANDOM | PICS - FUNNY - GAMING - ASKREDDIT - WORLDNEWS - NEWS - VIDEOS - IAMA - TODAYILEARNED - AWW - ADVICEANIMALS - SCIENCE - MUSIC - MOVIES - BESTOF - BOOKS - EARTHPORN - EXPLAINLIKEIM
reddit
hot
new
rising
controversial
top
gilded
wiki
promoted
subscribed
your front page
explore
multireddits
redditnews
redditpets
multibeta
explore multis
create
everything
from all subreddits
saved

new! create sets of subreddits to view together.
for starters, try one of these:
to hide these samples, create a multi of your own:

trending subreddits /r/OutOfTheLoop /r/BoredCelebs /r/ukraina /r/Glitch_in_the_Matrix /r/StarWars 23 comments

1 1867 Everyone, meet my little brother (imgur.com)
submitted 2 hours ago by Slothduction to /r/AdviceAnimals
71 comments share save hide report

2 3929 This duck sucks. (i.imgur.com)
submitted 6 hours ago by Bodysalt to /r/funny
588 comments share save hide report

3 3235 Amazing what mods can do for a 10 year old game (HL2) (i.imgur.com)
submitted 5 hours ago by DrSlappyPants to /r/gaming
955 comments share save hide report

4 2735 Early Movie Concept Arts (imgur.com)
submitted 5 hours ago by chaths to /r/movies
640 comments share save hide report

5 3854 GOOD BOY (i.imgur.com)
submitted 7 hours ago by CORNLORD to /r/gifs
644 comments share save hide report

6 3096 This year's birthday treasure hunt for my son. (imgur.com)
submitted 7 hours ago by crujones43 to /r/pics
467 comments share save hide report

7 2330 John Oliver reads an unexpected and hilarious response from POM Wonderful after skewering them on his show
submitted 5 hours ago by theruins to /r/videos
188 comments share save hide report

8 2990 While I was drinking coffee this morning, my dog brought me this. I have no idea where it came from. (i.imgur.com)
submitted 7 hours ago by NikonNights to /r/aww
509 comments share save hide report

9 2062 /r/worldnews is currently under a downvote attack - here's what you need to know, and what you can do (self.worldnews)
submitted 4 hours ago by slapchopsuey to /r/worldnews
794 comments share save hide report

10 2559 TIL researchers have invented a helmet that gives humans simulated 360° of Vision, allowing a person to avoid object tossed outside the normal human field of view. (irisa.fr)
submitted 6 hours ago by hans_grouper to /r/todayilearned
291 comments share save hide report

The main post discussing the down-vote attack on /r/worldnews was voted onto the front page of the site showing the most popular post across the entire Reddit

All through the experiment, long comment threads of speculation ensued as to what our sock puppet were doing, especially about their perceived agenda. A memorable accusation called our scripts “Putin-bot” based on the (incorrect) observation that only the pro-Russian news posts were not down-voted.

The same experiment ran concurrently on /r/netsec where the response there was markedly different. The /r/netsec moderators put up a post frankly acknowledging that apart from styling /r/netsec to hide down-votes altogether, they were powerless do anything else besides informing the Reddit administrators. In the comment thread below the post, Reddit administrators suggested quite strongly that they not have a handle on this simple attack. When raising this issue, the /r/netsec moderator quotes from the Reddit community manager as saying: [49]

> I don't know what else to tell you...Any site you go to will have problems similar to this, there is no ideal solution for this or other problems that run rampant on social websites.. if there was, no site would have any problems with spam or artificial popularity of posts.

Another admin commented that:

> You had a group of about 20 bots that were being used to down-vote posts in the subreddit. We rendered downvoting from those accounts ineffective, but to make it more difficult for the controller of the bots to realize that they’ve been disabled, we still need to make it look like their votes are applying. If we just throw away their votes entirely, the controller’s going to see that their bots have been blocked, and change up what they’re doing immediately.

What is particularly surprising was that their count of our sock puppets was considerably off: at the time we were using 50 sock puppet accounts to mass down-vote and they had only identified 20. The sole remediation done was to flag as spam the limited sock puppet votes which was crude and ineffective. Since the administrators seemed content to wait it out as well, we called off our sock puppets.

Trickle down-voting

To down-vote without causing this uproar we tried a variation — “trickle down-voting”. The mass down-vote strategy threw every sock puppet at each new link, keeping its score well below zero, but this was noticeable. A refinement of the technique was to keep the score as close to zero as possible; this meant that each new post’s score had to be monitored and as soon as up-votes were detected, our sock puppets were dispatch to counteract it with a down-vote.

We ran this for several days without users noticing, and was considered a success.

The overall effect of trickle down-voting was to lower the scores of posts that make it onto a subreddit front page. An obvious use of this is:

1. Keep a subreddit entry score to the front page low through trickle down-voting,
2. Identify candidate posts for promotion and then add up-votes, effectively allowing the piece to slingshot onto the front page.

As an aside, although Reddit tried hard to prevent users from seeing the actual vote scores of a post, we were able to discover a method of divining a story’s score. We were able to create an oracle out of the user preference that sets the users score threshold (posts below the threshold are not visible). By repeatedly changing this threshold, we can determine whether a post’s score was above or below a certain number and were able to narrow this down to an exact score.

Discussion

Promoting individual posts with up-votes is possible, particularly with new posts. When suppressing a post with down-votes, it is best to start as soon as possible after a post is created, to have a good chance of dragging the post score below the user-visibility threshold as soon as possible. Once out of sight, the post is much less likely to attract organic up-votes.

The surprising effect of the straight-forward down-voting attack can be very useful. It is fairly clear from the 1700 comments our actions generated that a number of redditors spent a great deal of that day engaged in discussing our attack. Had we triggered this ham-fisted attack while news of a new BP oil spill was breaking,

[49] http://www.reddit.com/r/netsec/comments/24w5l7/attempted_vote_gaming_on_rnetsec/

BP would have been grateful indeed for the number of users sucked into exposing lizard people and Putin-bots (instead of discussing and sharing details on the spill). It was exemplary misdirection.

The speculation surrounding the down-voting attack suggests another use. Clumsy or obvious sock puppets make for a good False Flag operation. A series of clumsy attacks can be used around a specific topic and then waiting for it to be detected (or simply exposing it with another sock puppet ourselves). The natural implication would be that the attacks are being carried out against the topic. In line with the earlier mentioned Streisand Effect we would expect to see increased interest around the topic as a result.

More generally the moderator and administrator comments in response to the detected obvious attacks exposed the lack of powerful or effective tools for dealing with straightforward disturbances. We will certainly be revisiting them later in the project when building tools for detection and prevention. The clumsy attacks were easy to spot as the sock puppets had had many similar characteristics:

- they all voted in sync,
- the signup/registration times were close together,
- all the sock puppets were operating from publicly known open proxies,
- the browser headers were all very similar (including user agents) and very different to that of normal browsers, and
- the sock puppets had low karma due to them having little interaction with any other accounts.

Moderators, however, would have had little chance of spotting these given how little user information is available to them in the moderation panels. This is clearly an area for potential improvement by Reddit's developers.

Two months after our down-vote attack, Reddit changed the site to hide down-votes by default. Interestingly, the score is only hidden when viewing the site in a browser; the vote numbers are still available from the API endpoints. This however does not solve the problem of down-vote attacks, it only stops most people from being able to notice it and is, therefore, a step backwards.

It is clear that the reddit back-end does take some steps to detect gaming, but it is just as clear that humans involved in administration and moderation of subreddits have little insight into the process. The problem with this approach is that it requires blind trust that Reddit is effectively defeating everything from spam to sock puppetry, without any way to evaluate their capability for this or having much input in the process. A better approach would be to make both more data available to subreddit moderators and having better tools on hand to quickly spot and stop potential sock puppet subterfuge.

HackerNews and Karma Growth

HackerNews is a site for the submitting, voting, and commenting on posts and links. It is similar to Reddit. except its audience is general highly technical and is focused on startups. Both HackerNews and Reddit automatically score users based on their contributions to the platform (although differently).

Aims

On HackerNews, users earn "karma" by posting links and comments. The net result of up-votes minus down-votes on these submissions determines users' karma score. The aim here is to demonstrably build karma on a user or bot.

Approach

We had a legitimate HackerNews user profile that posted 22 thoughtful submissions over a lengthy period just shy of three years. These well-curated posts covered a variety of topics from tools built, to conference proceedings shared to even op-ed pieces from major news houses.

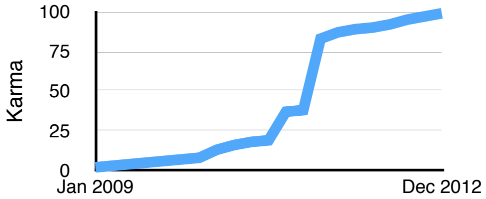
Clearly other HackerNews users did not find the content as appealing as we did, as the user had a karma score of only 99 in all that time. At 99 points, our user had not amassed enough "karma" to be taken seriously despite a fair bit of thoughtful curation at human timescales. As an experiment, we swapped careful curation at human timescales, for predictable selection at machine timescales.

A script was built to monitor a handful of popular blogs for new posts, and would immediately submit them to HN on spotting them.

Results

The result was a stark success. Over the initial 35 month period of manually curated posts the user's karma growth peaked at 99.

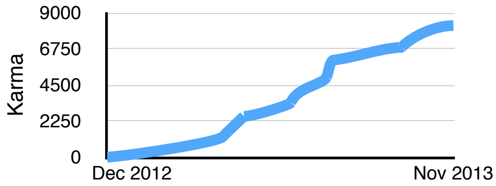
Once we started posting from the popular sites, the karma skyrocketed over the next 10 months from 99 to around 8000.



Karma grown over the initial manual posting period

Discussion

Clearly the automated building of karma is possible and this approach could apply equally well to Reddit. Combined with manual user commenting and a bit of patience, this could yield several high karma users over a relatively short period. This seriously calls into question the value that is placed on high karma as a means of assessing how trustworthy a user is on sites where karma is driven by votes on links.



Karma growth showing steep rise after changing to automatic posting

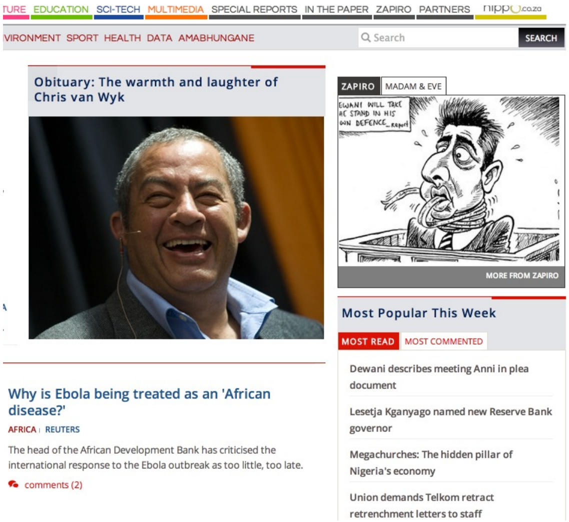
News Sites

News organizations' websites are attractive targets for trying to manipulate public opinion. Indeed, they exist to inform and shape public opinion and the key challenge here is to identify areas that our sock puppets can have measurable effects on.

Aims

News sites very often feature a panel on their front page that shows the current most popular articles on the site. These panels are often quite prominent on the front page, so users can be enticed into clicking on the page. This creates an interesting knock-on effect. Once an article is raised to this level of prominence (usually because it has been well read) it is now given the top spot, where it is more likely to gain even more views.

Our aim here was to influence what shows up on the panel, thereby drawing attention to stories we wish to promote and diverting attention from stories we consider unfavorable by supplanting their position with our own stories. Fortunately, we are easily able to determine the success of our attacks in this case.



Obituary: The warmth and laughter of Chris van Wyk

Why is Ebola being treated as an 'African disease?'

AFRICA | REUTERS

The head of the African Development Bank has criticised the international response to the Ebola outbreak as too little, too late.

Most Popular This Week

Dewani describes meeting Anni in plea document

Lesetja Kganyago named new Reserve Bank governor

Megachurches: The hidden pillar of Nigeria's economy

Union demands Telkom retract retrenchment letters to staff

The front page of the Mail & Guardian news site showing the most popular panel

Approach

We looked at a number of news sites and examined how the "popular" panel was generated, and it turns out it is not always just a hit counter. The sites that ultimately were tested were selected due to their use of alternative metrics. The approach in general was to examine how much influence could be brought to bear on the panel. Our experiment therefore was to promote articles into these "popular" sections.

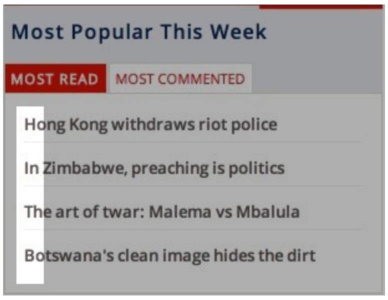
Results

Each news site attacked is discussed separately below.

Mail & Guardian

The Mail & Guardian is a South African weekly newspaper. The front page of the site featured a "Most Popular This Week" panel, and default ordering shown is "Most Read". This ranking was determined by page views which we could easily influence by just requesting the page we wished to promote.

We were able to get an article onto the front page with around 20 000 page requests, and were able to move an article to the top of the panel with a further 4000 page requests. These were all done from a single IP address. This allowed us complete control over which articles appeared in the panel (and in which order). We spelt "PUNS" and then later "HITB" to demonstrate this. In other words, we fully controlled the stories and their order.



Most Popular This Week

MOST READ MOST COMMENTED

Hong Kong withdraws riot police

In Zimbabwe, preaching is politics

The art of twar: Malema vs Mbalula

Botswana's clean image hides the dirt

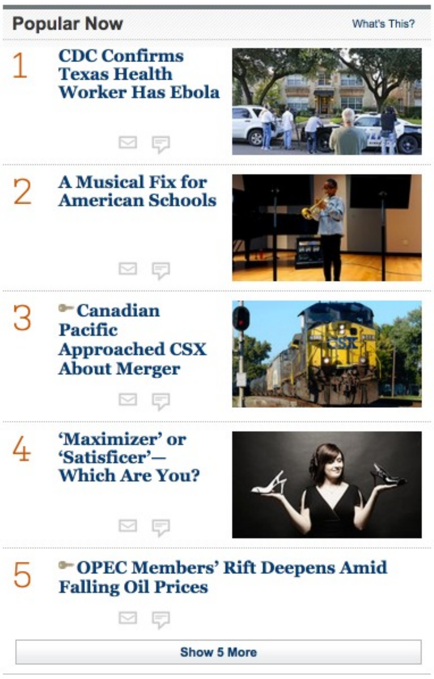
Our rearrangement of the articles on the Mail & Guardian panel to read "HITB"

Wall Street Journal

The Wall Street Journal's "Popular Now" panel ranked the articles by a weighted combination of several measures of each items' popularity. The ranking score comprised of page views (30%), Facebook shares, (20%), Twitter shares (20%), email shares (20%) and comments (10%).

By influencing page views and the Twitter share count of a link, we could influence the final ranking with a 50% weight. (How exactly the measures were combined to get the final rank was not specified, but we do not need to know how this was done, just that we were able to influence it enough to have a visible, deterministic effect.)

The Twitter API reported the number of times a link had been shared. The interesting thing about this count, was that it increased every time a user tweeted the link, even if the user had already tweeted the link previously. The only restrictions were that the same user could not Tweet the exact the same text in a Tweet more than once, and that the users tweeting was rate limited. To get around this we modified tweets with an incrementing counter, and had 7 Twitter accounts tweeting our link. (We operated slowly enough to avoid hitting the rate limit.)



Popular Now

1 CDC Confirms Texas Health Worker Has Ebola

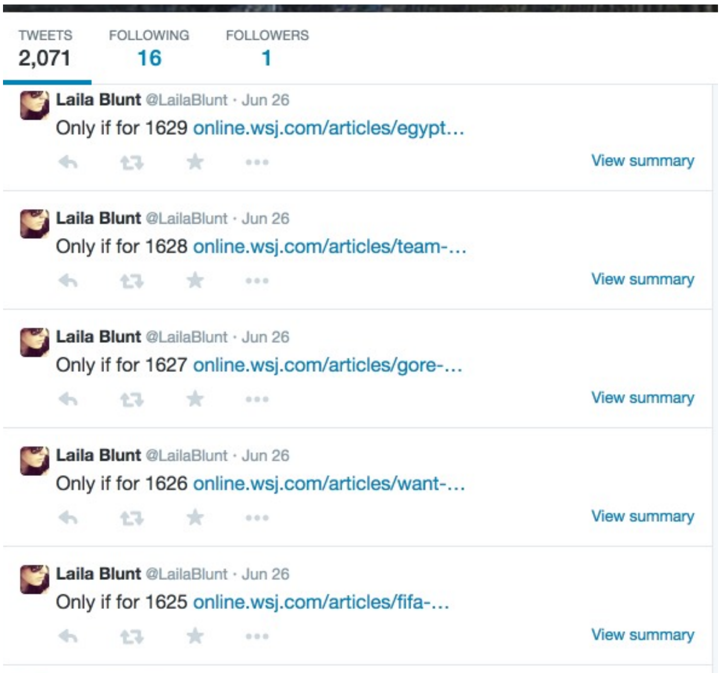
2 A Musical Fix for American Schools

3 Canadian Pacific Approached CSX About Merger

4 'Maximizer' or 'Satisficer'— Which Are You?

5 OPEC Members' Rift Deepens Amid Falling Oil Prices

The most popular panel on the front page of the Wall Street Journal



TWEETS 2,071 FOLLOWING 16 FOLLOWERS 1

Laila Blunt @LailaBlunt · Jun 26
Only if for 1629 online.wsj.com/articles/egypt...

Laila Blunt @LailaBlunt · Jun 26
Only if for 1628 online.wsj.com/articles/team-...

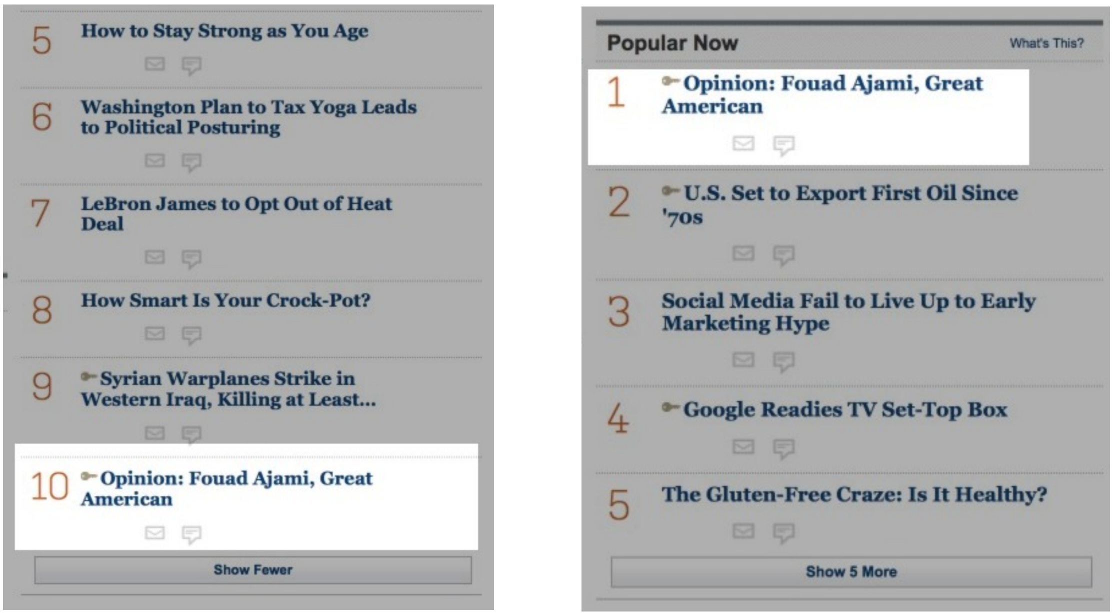
Laila Blunt @LailaBlunt · Jun 26
Only if for 1627 online.wsj.com/articles/gore-...

Laila Blunt @LailaBlunt · Jun 26
Only if for 1626 online.wsj.com/articles/want-...

Laila Blunt @LailaBlunt · Jun 26
Only if for 1625 online.wsj.com/articles/fifa-...

The same user repeatedly tweeting Wall Street Journal links to bump up their Twitter link share count

The combined effect of the tweeting and page requests moved an article from 10th to the 1st position on the list in under 7 hours. Interestingly, it did not matter that the article was paywalled and not actually available to us, it could be promoted without needing to view the article. Generating tweets and page requests was also used to launch an unlisted article onto the top 10 list within 40 minutes.



5 How to Stay Strong as You Age

6 Washington Plan to Tax Yoga Leads to Political Posturing

7 LeBron James to Opt Out of Heat Deal

8 How Smart Is Your Crock-Pot?

9 Syrian Warplanes Strike in Western Iraq, Killing at Least...

10 Opinion: Fouad Ajami, Great American

Show Fewer

Popular Now What's This?

1 Opinion: Fouad Ajami, Great American

2 U.S. Set to Export First Oil Since '70s

3 Social Media Fail to Live Up to Early Marketing Hype

4 Google Readies TV Set-Top Box

5 The Gluten-Free Craze: Is It Healthy?

Show 5 More

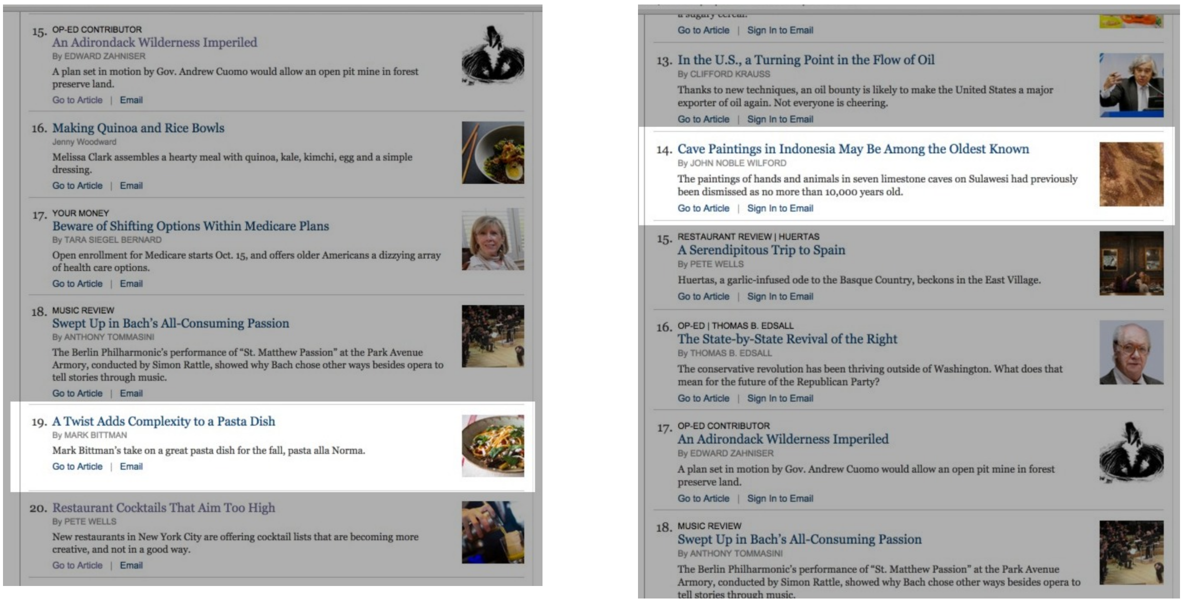
Moving an article from the bottom to the top of the Wall Street Journal Panel

The New York Times

The default view of the New York Times' "most popular" panel was ranked by the number of times an article was emailed over the past 24 hours. Each article has a link allowing a user to "share it" via email. However, sharing via emails is only available to registered user accounts.

Automating account creation is straightforward. The sign up requires an email and password, and the email address is never verified. At the time of writing we control over 30 000 accounts registered on the site, all using unique email addresses of disposable mail services. Signing up 30 000 accounts took about 3 hours to complete.

With these 30 000 accounts we were in a position to share articles via email at scale. The effect of the email shares was to reliably promote an article on the list from 20th position in the list to 15th or 14th. Completing all the email shares takes about 3 hours, after which is there is delay before the final result is seen.



15. OP-ED CONTRIBUTOR
An Adirondack Wilderness Imperiled
By EDWARD ZAHNISER
A plan set in motion by Gov. Andrew Cuomo would allow an open pit mine in forest preserve land.
Go to Article | Email

16. Making Quinoa and Rice Bowls
Jenny Woodward
Melissa Clark assembles a hearty meal with quinoa, kale, kimchi, egg and a simple dressing.
Go to Article | Email

17. YOUR MONEY
Beware of Shifting Options Within Medicare Plans
By TARA SIEGEL BERNARD
Open enrollment for Medicare starts Oct. 15, and offers older Americans a dizzying array of health care options.
Go to Article | Email

18. MUSIC REVIEW
Swept Up in Bach's All-Consuming Passion
By ANTHONY TOMMASINI
The Berlin Philharmonic's performance of "St. Matthew Passion" at the Park Avenue Armory, conducted by Simon Rattle, showed why Bach chose other ways besides opera to tell stories through music.
Go to Article | Email

19. A Twist Adds Complexity to a Pasta Dish
By MARK BITTMAN
Mark Bittman's take on a great pasta dish for the fall, pasta alla Norma.
Go to Article | Email

20. Restaurant Cocktails That Aim Too High
By PETE WELLS
New restaurants in New York City are offering cocktail lists that are becoming more creative, and not in a good way.
Go to Article | Email

a sugary cereal.
Go to Article | Sign In to Email

13. In the U.S., a Turning Point in the Flow of Oil
By CLIFFORD KRAUSS
Thanks to new techniques, an oil bounty is likely to make the United States a major exporter of oil again. Not everyone is cheering.
Go to Article | Sign In to Email

14. Cave Paintings in Indonesia May Be Among the Oldest Known
By JOHN NOBLE WILFORD
The paintings of hands and animals in seven limestone caves on Sulawesi had previously been dismissed as no more than 10,000 years old.
Go to Article | Sign In to Email

15. RESTAURANT REVIEW | HUERTAS
A Serendipitous Trip to Spain
By PETE WELLS
Huertas, a garlic-infused ode to the Basque Country, beckons in the East Village.
Go to Article | Sign In to Email

16. OP-ED | THOMAS B. EDSALL
The State-by-State Revival of the Right
By THOMAS B. EDSALL
The conservative revolution has been thriving outside of Washington. What does that mean for the future of the Republican Party?
Go to Article | Sign In to Email

17. OP-ED CONTRIBUTOR
An Adirondack Wilderness Imperiled
By EDWARD ZAHNISER
A plan set in motion by Gov. Andrew Cuomo would allow an open pit mine in forest preserve land.
Go to Article | Sign In to Email

18. MUSIC REVIEW
Swept Up in Bach's All-Consuming Passion
By ANTHONY TOMMASINI
The Berlin Philharmonic's performance of "St. Matthew Passion" at the Park Avenue Armory, conducted by Simon Rattle, showed why Bach chose other ways besides opera to tell stories through music.

Moving an article up the New York Times most popular list

The only limitation here was that each IP address was limited to 10 000 email shares regardless of the user account doing the sharing. This is hardly a limiting factor at all, as the email shares were sent from Amazon EC2 machines for which a reboot gets the machine a new IP address. It is also concerning that the NYT email sharing server was made to send 30 000 emails several times during our experiments without raising any serious flags.

Discussion

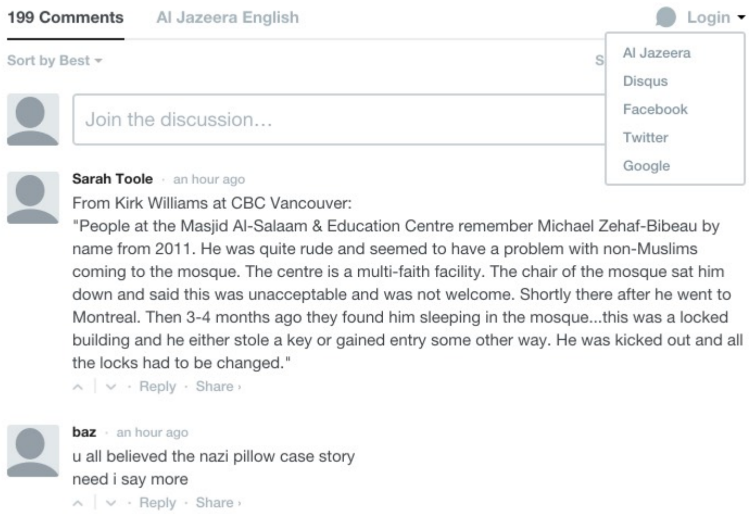
It certainly seems that these panels were not designed to be robust against external manipulation. In cases where the ranking on the panel is dependent on a third party service (like Twitter), the panel is now exposed to whatever quirks are present on that service. This is well demonstrated on the Wall Street Journal by how easy it is to increase the Twitter share count of a link with a single user.

Some manipulation can be made a bit more difficult. Adding a CAPTCHA and email verification to the New York Times account sign up would increase the cost of account creation. However, on measures such as page views, perhaps IP limits needs to be considered.

Given the prime positions these panels occupy on news sites and the ease of manipulation, they deserve a more attention than they currently receive.

Comment Hosting Services

Disqus and Livefyre are two major comment hosting services and both provide a drop-in commenting system that web site owners can easily insert into their pages. Disqus processes 20 million comments a month and features on everything from major new sites like Al Jazeera, and CNN, all the way down to small Wordpress blogs and Tumblrs. Livefyre is present across a similar range: their comments system is used on Fox News, Sky News, and many small blogs. The attraction of these commenting services is both their ubiquity across the web in general, but even more importantly across the pages of major news sites.



Sarah Toole · an hour ago

From Kirk Williams at CBC Vancouver:
"People at the Masjid Al-Salaam & Education Centre remember Michael Zehaf-Bibeau by name from 2011. He was quite rude and seemed to have a problem with non-Muslims coming to the mosque. The centre is a multi-faith facility. The chair of the mosque sat him down and said this was unacceptable and was not welcome. Shortly there after he went to Montreal. Then 3-4 months ago they found him sleeping in the mosque...this was a locked building and he either stole a key or gained entry some other way. He was kicked out and all the locks had to be changed."

baz · an hour ago

u all believed the nazi pillow case story
need i say more

Screenshot of a typical Disqus comment widget

The pervasiveness of these comment systems is partly explained by how easy it is for site owners to adopt them. All the site owner does is embed some HTML and Javascript in the page they serve their users, and comment widgets render on the pages providing a fully-fledged commenting system. Typically there is also an administrative back-end for managing the comments posted to a site.

Aims

Since the comment widget covers feature on many well-read pages on the web our two main aims here are:

- obtaining sock puppet accounts, and
- influencing what gets shown by default on the comment widgets.

Given that these same commenting systems are so ubiquitous, we were primarily interested in sock puppetry techniques that work regardless of which site the commenting system is on.

Approach

Our approach to the commenting systems is to test the possibilities of:

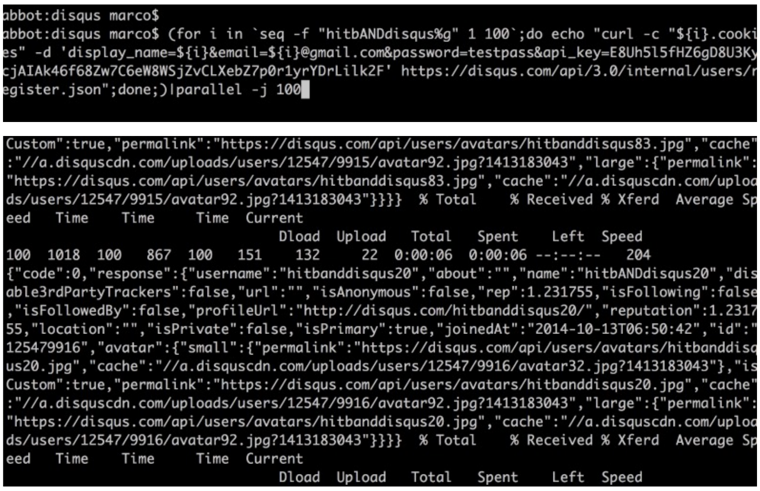
- creating sock puppet accounts *en masse*,
- impersonating user accounts to perform actions on their behalf,

- promoting a comment to the top of the list of comments shown by default,
- suppressing a comment from its position on the list of comments shown by default,
- removing a comment entirely,

It is simple to verify whether any of these are achieved, so in the Results section we will note which tests succeeded and what strategy worked.

Results

Sock puppetry on both these platforms is not just possible, it is easy and fun!



```
abbot:disqus marco$
abbot:disqus marco$ (for i in `seq -f "hitbANDdisqus%g" 1 100`;do echo "curl -c "${i}.cooki
es" -d 'display_name=${i}&email=${i}@gmail.com&password=testpass&api_key=E8Uh5l5fHZ6gD8U3Ky
cjAIAk46f68Zw7C6eW8WSjZvCLXebZ7p0r1yrYDrLilk2F' https://disqus.com/api/3.0/internal/users/r
egister.json";done;)|parallel -j 100
```

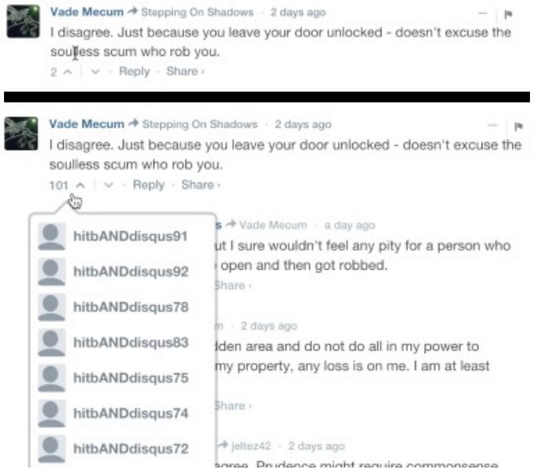
```
Custom":true,"permalink":"https://disqus.com/api/users/avatars/hitbanddisqus83.jpg","cache"
:"//a.disquscdn.com/uploads/users/12547/9915/avatar92.jpg?1413183043","large":{"permalink":
"https://disqus.com/api/users/avatars/hitbanddisqus83.jpg","cache":"//a.disquscdn.com/uploa
ds/users/12547/9915/avatar92.jpg?1413183043"}}}}  % Total    % Received % Xferd  Average Sp
eed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
100  1018  100   867  100   151    132     22  0:00:06  0:00:06 --:--:--   204
{"code":0,"response":{"username":"hitbanddisqus20","about":"","name":"hitbANDdisqus20","dis
able3rdPartyTrackers":false,"url":"","isAnonymous":false,"rep":1.231755,"isFollowing":false
,"isFollowedBy":false,"profileUrl":"http://disqus.com/hitbanddisqus20/","reputation":1.2317
55,"location":"","isPrivate":false,"isPrimary":true,"joinedAt":"2014-10-13T06:50:42","id":"
125479916","avatar":{"small":{"permalink":"https://disqus.com/api/users/avatars/hitbanddisq
us20.jpg","cache":"//a.disquscdn.com/uploads/users/12547/9916/avatar32.jpg?1413183043"},"is
Custom":true,"permalink":"https://disqus.com/api/users/avatars/hitbanddisqus20.jpg","cache"
:"//a.disquscdn.com/uploads/users/12547/9916/avatar92.jpg?1413183043","large":{"permalink":
"https://disqus.com/api/users/avatars/hitbanddisqus20.jpg","cache":"//a.disquscdn.com/uploa
ds/users/12547/9916/avatar92.jpg?1413183043"}}}}  % Total    % Received % Xferd  Average Sp
eed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
```

Screenshots of running the one line command to create 100 Disqus users in seconds

Disqus

Creating accounts on Disqus was straightforward. There were options to signup using social media accounts, but we signed up with email addresses. (The email addresses did not need to exist because there was no email verification.) As there was no CAPTCHA or any other anti-automation protection, this process was trivially automated and using a few lines of scripting on a single AWS instance (with a single IP), were able to register new Disqus accounts at a rate of about 1500 accounts per minute. These registered accounts could be used on any site which uses Disqus comments, instead of having to create a new batch of accounts per site.

Disqus allowed comments to be up-voted or down-voted, which affected their score. By default the comments widget shows the comments ordered by the those with the best score (up-votes less down-votes). To down-vote or post a comment a user has to be logged in, but up-votes can be cast by guest users.



Vade Mecum → Stepping On Shadows · 2 days ago

I disagree. Just because you leave your door unlocked - doesn't excuse the soulless scum who rob you.

Vade Mecum → Stepping On Shadows · 2 days ago

I disagree. Just because you leave your door unlocked - doesn't excuse the soulless scum who rob you.

101
Reply · Share ›
hitbANDdisqus91
hitbANDdisqus92
hitbANDdisqus78
hitbANDdisqus83
hitbANDdisqus75
hitbANDdisqus74
hitbANDdisqus72
s → Vade Mecum · a day ago
ut I sure wouldn't feel any pity for a person who
open and then got robbed.
Share ›
n · 2 days ago
lden area and do not do all in my power to
my property, any loss is on me. I am at least
Share ›
→ jeltez42 · 2 days ago
agree. Prudence might require commonsense

Before (top) and after (bottom) voting on a Disqus comment with our sock puppets

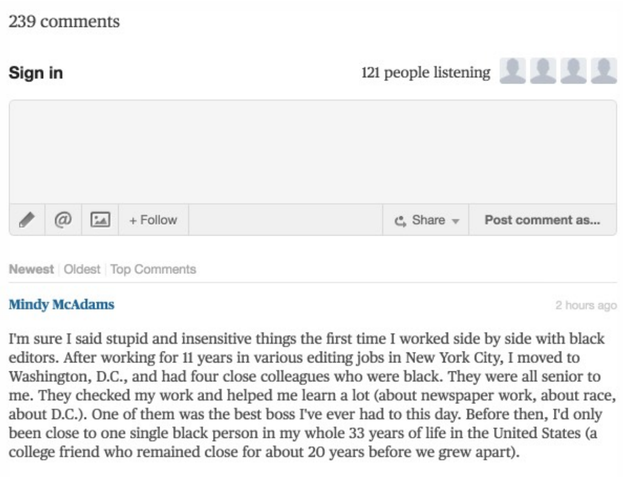
With our sock puppets we could easily up-vote comments we wanted to promote. The same effect could be achieved by up-voting as a guest voter, however since only one anonymous vote can be cast from a single IP, multiple source IP addresses are required.

Similarly we could down-vote a comment to lower its position in the default view. The only difference was that guest voters could not down-vote, meaning registered accounts had to be used.

Removing a comment from view was possible by flagging the comment as spam several times. (Once that comment was explicitly approved however, the process could not be repeated.)

Livefyre

We tested Livefyre Community Comments which offered a similar service to Disqus, and was also widely used. The comment widget by default ranked comments by how new they were. (There was also “Top Comments” for most liked.)



Mindy McAdams · 2 hours ago

I'm sure I said stupid and insensitive things the first time I worked side by side with black editors. After working for 11 years in various editing jobs in New York City, I moved to Washington, D.C., and had four close colleagues who were black. They were all senior to me. They checked my work and helped me learn a lot (about newspaper work, about race, about D.C.). One of them was the best boss I've ever had to this day. Before then, I'd only been close to one single black person in my whole 33 years of life in the United States (a college friend who remained close for about 20 years before we grew apart).

A typical Livefyre comment widget

Creating accounts was straightforward. There were options to sign in with other social media accounts, but we (once again) made use of the email signup. Similar to Disqus this registration was easily automated, as non-existing email addresses could be used, and there were no CAPTCHAs. We did not create many accounts, because they were not needed for the further tests.

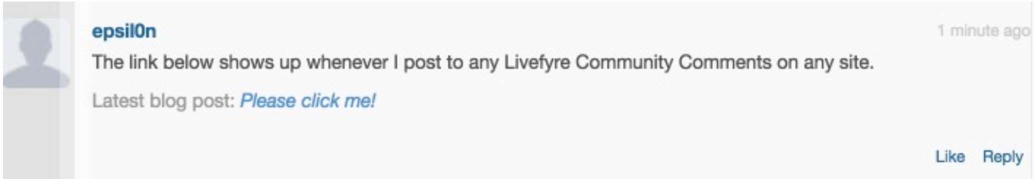
During the course of our research, we discovered a vulnerability on Livefyre comments that allowed an attacker to impersonate other users. Carrying out the attack was relatively simple: a victim user visiting a specially crafted page while logged into Livefyre comments on any other site had their Livefyre account hijacked by the malicious page. The crafted page contains a comment-widget for a site we control, which allowed the page to steal the user's token for performing Livefyre actions.

The attack worked as follows:

1. A logged in user visits our specially crafted page.
2. The comment widget for our own site renders on the page.
3. The visiting user is then added as an administrator to our site's comment system.
4. The comment widget is refreshed, and now includes the victim's token in the comment widget.
5. The malicious page is then able to steal the user's Livefyre token.

Once the user's Livefyre token is obtained, the attacker can post comments to any other site (or "like" comments) as that user, without the user being notified. The vulnerability has been reported to Livefyre, and they working are on a fix.

It is not difficult to distribute a link to other logged in users and spammers on Livefyre comments embed links directly in their comments. A neater (alternative) way to achieve this is to use the fact that site administrators have a "Latest blog post" link attached by default to their posts. We simply administer our own site, add a link on it, and have the link sent out every time we comment.



epsil0n

1 minute ago

The link below shows up whenever I post to any Livefyre Community Comments on any site.

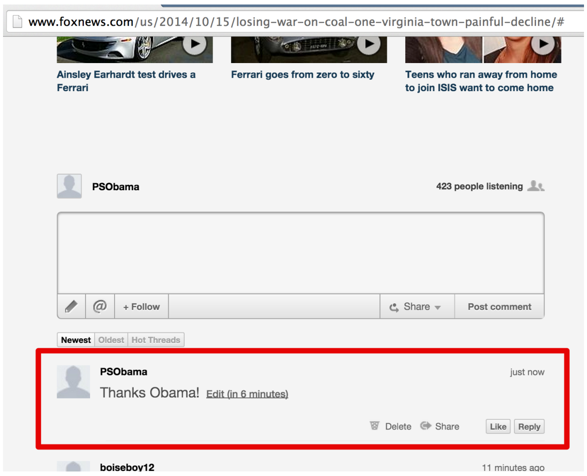
Latest blog post: *Please click me!*

A post showing our user's "Latest blog post" link which we control

As the default ordering of Livefyre comments is "newest first", simply posting a new comment was enough to feature a comment at the top. On a busy site like Fox News a post may stay at the top for a matter of minutes only due to its busy nature. To keep desired content at the top, our sock puppet accounts simply have to continuously post at a fast enough rate.

The same process of continuously posting new comments, will also knock undesired comments further down the list. Each new comment shifts all the older comments down by one on the default view of the comment widget.

To completely remove a comment from the list, flagging it as spam from 5 distinct accounts hid the comment from view. The number of spam flags required is configurable by the site administrator. This can be tested on the Fox News site without any programming by registering 5 accounts from a single IP, and flagging a comment as spam.



www.foxnews.com/us/2014/10/15/losing-war-on-coal-one-virginia-town-painful-decline/#

Ainsley Earhardt test drives a Ferrari

Ferrari goes from zero to sixty

Teens who ran away from home to join ISIS want to come home

PSObama

423 people listening

PSObama

just now

Thanks Obama! Edit (in 6 minutes)

boiseboy12

11 minutes ago

Posting a new comment on a Fox News article that pushes all the others down

Discussion

These two commenting systems were surprisingly easy to manipulate given the amount of space they occupy on many of the world's major news sites.

Both Disqus and Livefyre provided administrator interfaces for tracking user comment activity. On both comment systems site administrators could see the collection of IP addresses used by a commenter.

Tracking activity emanating from a single IP was not too far-fetched. This kind of data is useful to site administrators trying to spot sock puppetry in action.

Simply protecting account creation with email verification and CAPTCHAs would help slow down attackers trying to create many accounts. It seems that the comment hosts want to have as frictionless an experience as possible for users to sign up, so they can immediately beginning interacting on the site. However, as it is currently, this sets a very low bar for potential attackers.

Both comment systems could be configured to require moderator approval for all comments. Surprisingly, there is a non-trivial example of a comments system where this is done: The New York Times has their own comment system that requires moderator approval for all comments. Unsurprisingly, this needs enough staff to handle the moderation queue. Requiring moderator approval would certainly weed out the spam comments seen on some Livefyre deployments, and is likely to raise the bar for generating content for the comments, as seen with moderator approval on the Full Disclosure mailing list. However, moderation on content alone, is unlikely to be sufficient to effectively stop sock puppetry. More visibility on user activities however, should allow moderators and site admins to quickly spot nefarious activity.

Discussion and Future Work

This section gathers together general observations from the experiments run for Phase 1. These observations are from responses to creating sock puppets, running them, and finding ways to measure their effectiveness. The sections ends with consideration of future sock puppetry work.

The difficulty of creating multiple sock puppets accounts varied widely between platforms. The New York Times accounts and Disqus accounts were trivial to create, not even requiring a valid email address, while others like Twitter took more work. Simply adding CAPTCHAs and email verification to account creation raises the difficulty level for attackers. Extra protections may discourage new users signing up or participating, but it is good to note that the Al Jazeera (Arabic) poll still attracts a few thousand votes *despite* being protected by a CAPTCHA. A side benefit of using Google's reCAPTCHA for the CAPTCHA is that many open web proxies (which we used frequently for more IP addresses) are blocked by default. News site panels and polls that do not require any signing up, should have rudimentary protections (such as limiting heavily used IP addresses.) If the resulting metrics are inaccurate, it may be better to change the panel or poll rather than implying to users that the panel is representative of other users' views.

In a few cases, the platform in question relies on other sites to identify the user. However, where multiple options for other sites to identify the user exists, the defense against mass account creation is only as good as the weakest option. On Livefyre and Disqus we signed up using unverified email addresses, because it was completely trivial as opposed to creating multiple Facebook accounts. The Wall Street Journal's Most Popular panel relies on several counts of link popularity. Again, we went for the two easiest for us to influence (page views and Twitter shares), rather than the comments count which needed access to paywalled articles. Also, the other site may behave unexpectedly. In the WSJ case, it is both interesting and very useful that a single user can bump up the Twitter link share count just by repeatedly tweeting the same link.

It may be tempting to think that for cases where sock puppets require posting content, moderating all content would stop sock puppetry. However, our experiments on the Full Disclosure mailing list reveal otherwise. While this probably weeds out the poorest quality posts and spam, it did not did not stop us using sock puppets effectively. Even on commenting sites like Livefyre and Disqus, moderating all posts based on content alone seems unlikely to be enough to stop sock puppetry.

What could help instead is moderation that includes well-presented data about the users. Both Livefyre and Disqus include IP addresses of users and that will help spot a sock puppets that operate from the same IP address. There is data to back this view; a sock puppet army was uncovered on the CommonDreams site using IP information. However, beyond simple sock puppets current moderation tools are not of much help. More work could be done to make it easier for humans to spot suspicious sock puppet behavior.

Generally having some data available, particularly for moderators is useful for detecting and acting on sock puppetry. On Reddit, the site administrators have taken the opposite approach, which is to say they limit the data exposed to users. The moderators cannot see or do much about sock puppets mass down-voting on their subreddits. The limited visibility available to the moderators and users requires trusting that the site and site administrators have completely stopped all sock puppet attacks. Given that there is little data even about the actual vote scores (and now the hidden down-votes), users would be unable to know if, for example, mass down-voting was underway.

In general wherever sock puppetry works, especially if it does so poorly, these could be used for false flags, or for discrediting another user. One simple trick could be buying obviously fake followers for another Twitter user, and then publicly calling them out for it. Alternatively, to promote something, the Streisand Effect could be leveraged. Poor quality sock puppetry can be performed to clumsily suppress something, and then have it be discovered (or have another sock puppet point it out). Ideally, the "suppressed" content would be promoted by the discussions around and automatic resistance to the sock puppetry.

One of the our unexpected challenges was finding good tests to reliably measure the effects of our sock puppets. Future work should certainly look at the many other cases where sock puppets can have measurable effects, including attempting to explore untested possibilities on our Twitter experiments. On the other side, tools could be built to at the very least make it easy to spot the sock puppet attacks we carried out. This will be explored thoroughly in subsequent phases of this project.

Conclusion

The objectives of Phase 1 were to demonstrate the efficacy of sock puppets in shaping narratives and silencing opinions. In this report we have provided a comprehensive background to sock puppetry and demonstrated several techniques across a variety of platforms where this is measurably possible, and in some cases very easy. Phase 1 was successful and uncovered new techniques and approaches to build modern sock puppets.

Turning towards the next two phases of the project, it was apparent that the presence of user data shone a light on possible sock puppet activity, helping detect when sock puppets were present as opposed to keeping users and moderators in the dark. Subsequent phases of this project will examine detecting these attacks, and building tools to make this detection easier.

Through the conference talk, and paper we hope to raise awareness of the evolving threat of sock puppets. Along the way we accidentally caused an uproar on Reddit, played scrabble on the front page of a news site, managed to get our office IP address banned from a few sites, had many user accounts suspended and received over 100 000 emails from the New York Times. A solid day’s work.

Feedback on this research is most welcome. Please mail us with any thoughts, questions and comments at sockpuppets@thinkst.com.