

CentMail: Rate Limiting via Certified Micro-Donations

Sharad Goel

Jake Hofman

John Langford

David M. Pennock

Daniel M. Reeves

{goel, hofman, jl, pennockd, dreeves}@yahoo-inc.com

1. INTRODUCTION

The Internet has reduced the cost of communication to near zero, benefiting billions of people around the world. One consequence, however, is that unsolicited and unaccountable commercial communication, or *spam*, is also sent indiscriminately in massive quantities at low cost, imposing a large burden on recipients and on systems. Spammers have infiltrated nearly every form of online communication, including email, instant messaging, blog comments/trackbacks, and web pages/links. We propose a system for rate limiting Internet communications broadly, emphasizing the case of email.

Domain and content filtering are currently the first line of defense against spam. But domain filtering is difficult to apply when spammers send from legitimate domains (e.g., by opening email accounts at Yahoo! and Gmail). It also places an onerous burden on new domains to establish themselves as legitimate senders. Content filtering requires considerable effort to maintain as spammers constantly evolve to circumvent the latest filters.

Many people, most notably Bill Gates, have observed that adding a modest cost to sending email, for example by requiring postage stamps like ordinary mail, could significantly deter spam. Researchers have proposed and analyzed several such systems, including variations where the recipient keeps the payment, the recipient has the option of either keeping or refunding the payment [7], the sender “burns” human time or CPU cycles [1, 5], or the sender pays to a charity of their or the recipient’s choice [3].

Although an equilibrium where senders and receivers all adopt email stamps benefits nearly everyone, there is a serious *flag day*¹ problem, or coordination failure, that makes the equilibrium hard to reach from the status quo. Senders do not want to spend money buying stamps if recipients are not checking stamps, and recipients would not bother to check stamps if few senders use them. The hurdle for senders is heightened by the very real possibility that spammers who already hijack other people’s computers may now in addition drain the users’ stamp accounts of money to send spam or, worse, to funnel the money to themselves. It seems that the prospect of some day reducing spam is not enough to convince a critical mass of both senders and recipients to adopt a new protocol and monetary accounting infrastructure.

¹[http://en.wikipedia.org/wiki/Flag_day_\(software\)](http://en.wikipedia.org/wiki/Flag_day_(software))

--

Alice certified this email by donating \$0.01 to Sierra Club
This donation was matched by Bob’s Widgets
Powered by CentMail.net -- Do good. Fight spam.

Figure 1: CentMail stamps appear as an email signature that promotes the sender’s cause.

2. CENTMAIL

CentMail is a general economic framework for rate-limiting that directly addresses the issue of incentives, providing tangible net benefit to even the earliest adopters with no need for coordination. This tailoring of incentives is threefold. First, users receive stamps in exchange for making donations to charitable organizations of their choice. In this way, many users would incur no added financial burden since they already make these donations regardless of their participation in CentMail. In fact, 89% of U.S. households already make annual donations, with an average household contribution of \$1620 (or 3.1% of income) [2] and a median on the order of \$100. (A donation of \$100 yields enough stamps to send email to 10,000 people—27 recipients per day, every day of the year.) Second, the stamps themselves are implemented as email signatures (see Figure 1) that promote both the sender’s cause and the sender’s support of that cause. A recent survey estimates that “people are 100 times more likely to donate when asked by a friend or family member than an anonymous solicitation” [4]. Third, users amplify their impact via matching donations, either by a corporate sponsor acknowledged in the message signature, or by the mail provider who may eventually see a reduction in spam-associated costs, estimated to be on the order of billions of dollars per year worldwide [6]. For many potential users, these design choices in sum yield net benefit, even in the absence of other participants in the system.

Although CentMail offers users benefits even in the absence of coordination, cooperation is still required for it to function as a spam deterrent. We note that in this regard, CentMail improves upon existing proposals in that CentMail stamps serve as advertisements for the system itself. This allows us to leverage the latent social network of email contacts to encourage adoption of the system. If enough senders join, recipients may take notice and begin to whitelist stamped email, allowing them to tune their content-based filters more aggressively, increasing the incentive for senders to stamp email, forming a virtuous cycle.

We do not expect to see complete adoption of CentMail.

However, even with limited adoption, we believe stamping to be an effective tool that works in conjunction with aggressive domain and content-based filtering to detect and deter spam.

A similar and independent effort at IBM Research in 2004 aimed to promote “charity seals” in email [3]. To our knowledge, that system was not implemented. Our main contribution is to make the idea concrete by defining a formal protocol, implementing a working prototype of the service, and analyzing the benefits and drawbacks of the approach and how we envision the service’s promotion and adoption.

2.1 The Protocol

The CentMail protocol supports authentication of both emails and arbitrary text documents. For example, a “document” could correspond to a comment on a weblog, or a listing of links on a web page. In each case, CentMail certifies that the content was validated by a charitable donation. The two key operations are certification and verification:

```
Centmail.certify(amount, digest)
return null
```

```
Centmail.verify(digest)
return {amount, queries}
```

These function calls are authenticated, and in particular, the user making the call is identified by the global parameter `Centmail.user`.

The `certify` function takes as input an `amount` to donate and the `digest`, or SHA-1 hash, of the content.² It is generally in the sender’s interest to append a *nonce* (i.e., a randomly generated string) to their content to ensure each message is unique—although this is not explicitly required by the protocol. When a message is certified, the CentMail server debits `CentMail.user` and stores the `digest` for later verification. In order to efficiently scale, message digests are eventually expired, and hence are maintained on the CentMail server only temporarily.

To verify content, the user passes the document `digest` to `Centmail.verify`. This call returns the `amount` which was donated, and `queries`, the number of times the content has been verified. The return value `queries` is a crucial piece of information since the certifier’s “payment” (i.e., donation) is less meaningful when the content is consumed by multiple individuals. For example, in the case of email, donating \$0.01 for an email which is ultimately sent to 1000 people is less of a commitment than donating \$0.01 for an email sent to a single individual.³ Typically, recipients would accept messages when `amount/queries` is at least \$0.01, and treat messages not meeting this threshold as effectively unstamped. In this latter case of unstamped—or effectively unstamped—messages, existing domain-based and content-based techniques could still be applied to classify email. The recipient, however, is free to enforce any filtering policy of their choice.

²In the case of a plain text document, computing the message hash is straightforward. For email, however, care must be taken so that determining which header fields to include, and their order, is unambiguous.

³Often one can avoid the problem of multiple recipients consuming the same content by using different nonces for each recipient. Then instead of sending 1000 people the “same” certified email message, for example, each recipient would in fact be verifying their own unique copy.

Aside from certifying messages, it is often useful for applications to rate limit requests (e.g., for account creation, or for posting comments to blogs). CAPTCHAs are typically used in these contexts, requiring users to burn “human cycles.” CentMail facilitates an alternative, economic approach to rate limiting which allows third parties to ask users to burn (i.e., donate) money. This feature is intended for web-based applications, and is implemented through an additional function call:

```
Centmail.request(amount, transID, returnUrl)
return null
```

When the requester makes this call, the end user (i.e., the individual being asked to make a donation) is redirected to the CentMail website to confirm the donation. Afterward, CentMail posts an authenticated response to `returnURL` and redirects the user back to the originating site. The requester receives confirmation that a donation was made, but no other identifiable information about the user.

2.2 The Implementation

A beta implementation of CentMail is available at `CentMail.net`. In addition to an initial implementation of the CentMail API on the server side, we have developed a CentMail plug-in for Thunderbird, the popular open source email client, a Firefox plug-in for web-based email services, including Yahoo! Mail and Gmail, an Apple Mail plug-in, and perl scripts for clients such as Pine, Mutt and Evolution that support filtering email through arbitrary scripts.

Full Version of the Paper

At `CentMail.net` we have available the full version of this paper, which details protocol specifics, proves correctness properties, discusses additional related work, and addresses concerns such as how to deal with mailing lists.

3. REFERENCES

- [1] A. Back. Hashcash—a denial of service counter-measure (5 years on). Tech Report, 2002.
- [2] M. S. Brown. *Giving USA: The Annual Report on Philanthropy for the Year 2007*. Giving USA Foundation, 2007.
- [3] P. Capek, B. Leiba, and M. N. Wegman. Charity begins at ... your mail program. <http://www.research.ibm.com/spam/papers/charity-seals.pdf>, 2004.
- [4] P. B Carroll. Charity cases, July 14, 2008. The Wall Street Journal, <http://online.wsj.com/article/SB121554292423936539.html>.
- [5] C. Dwork and M. Naor. Pricing via processing or combatting junk mail. In *Proceedings of CRYPTO’92*, pages 137–147, 1993.
- [6] D. Ferris, R. Jennings, and C. Williams. The Global Economic Impact of Spam, 2005. Technical Report 409, 2005. <http://www.ferris.com>.
- [7] Thede C. Loder, Marshall W. Van Alstyne, and Rick Walsh. An economic response to unsolicited communication. *Advances in Economic Analysis & Policy*, 6(1), 2006. Available at: <http://www.bepress.com/bejeap/advances/vol6/iss1/art2/>.

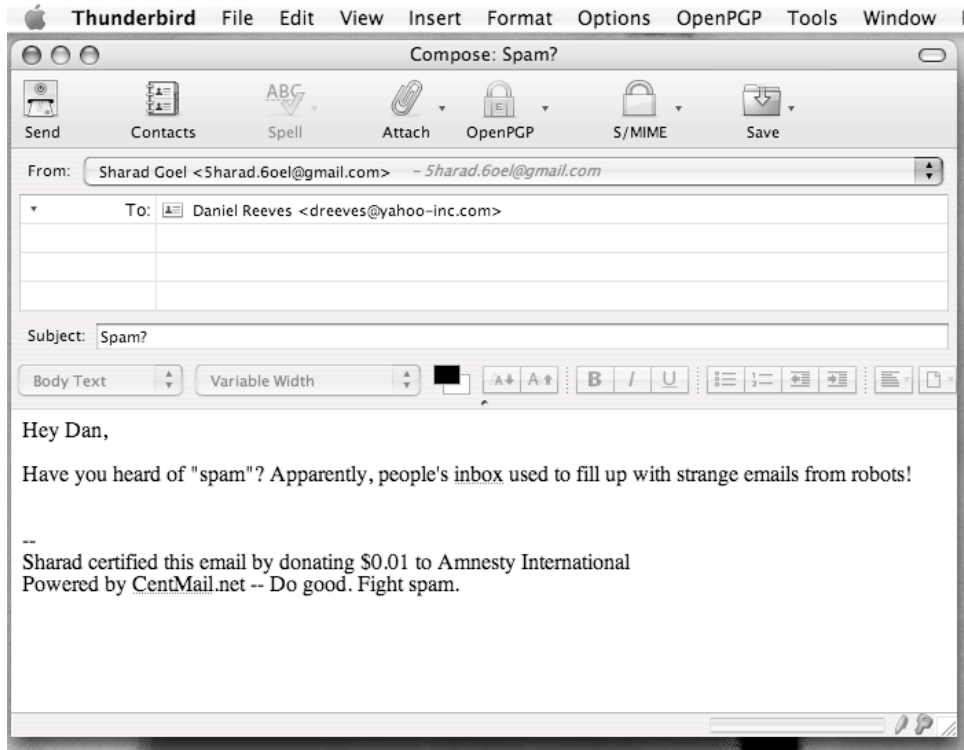


Figure 2: CentMail plug-in for the Thunderbird email client.

CentMail Do GOOD. FIGHT SPAM. Sign In

HOW IT WORKS

- 1. DONATE**
Donate \$5 to a participating charitable organization of your choice and receive 500 CentMail stamps.
- 2. STAMP**
Promote your favorite cause by attaching a CentMail stamp to each email you send.
- 3. VERIFY**
Anyone you email can automatically verify your donation and confirm you're not a spammer. Since spammers send millions of emails every day, it is prohibitively expensive for them to donate even just 1¢ per email.

GET STARTED

[Manage Account](#) [Create Account](#)

Learn more in our [technical paper](#).




Figure 3: <http://CentMail.net>