I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

DEBIN LIU*

The Economics of Proof-of-Work

Abstract: Proof-of-Work is a set of cryptographic mechanisms that increase the cost of initiating a connection. Currently recipients bear equal or greater cost per connection as initiators. This allocation of cost enables resource abuse, exhaustion attacks and denial of service, such as spam. The design goal of Proof-of-Work is to reverse the economics of connection initiation on the Internet. The first economic examination of Proof-of-Work, in the case of spam, argued that Proof-of-Work would not work. This result was based on the difference in email production costs between legitimate and criminal enterprises. This work illustrates that the difference between production costs does not remove the efficacy of Proof-of-Work when work requirements are weighted. This paper suggests that Proof-of-Work will work with a simple reputation system modeled on the systems currently used by commercial anti-spam vendors. This paper also addresses the idea that the proposed variation on Proof-of-Work changes the nature of the corresponding proofs from a token currency model to a notational currency model.

* Ph.D. Candidate Informatics Indiana University, Bloomington. First, I would like to thank Professor L. Jean Camp for her continuous support on this topic. Professor Camp was always there to listen and give advice. Second, I would like to thank WEIS 2006 reviewers for their comments. Partial financial support from the NET Institute (http://www.NETinst.org) for this research is gratefully acknowledged as well.

I. INTRODUCTION

Spam is a significant problem in that it consumes vast network and human resources. If the Internet is an attention span economy, then spam is wholesale theft. A secondary danger of spam is that it significantly reduces the value of email as a channel for customer-merchant and employer-employee communications.

Spam is so profitable that estimates of spam as a percentage of all email have increased even as the total volume of email increases. CipherTrust estimates more than 85% of all email messages are unwanted spam.[1] Spam is malicious network activity enabled by the otherwise virtuous cycle of network expansion. As the network expands, spam becomes more profitable and thus, increases. Spam is also a vector for other activities: distribution of malicious code, phishing attacks, and old-fashioned fraud.

The core challenge of defeating spam is that the sender bears almost no cost to send email. The cost is borne by the network service providers and the recipients. In order to solve this problem, Proof-of-Work can be used to alter the economics of spam by requiring that the sender commit to a per-email cost.[2] Proof-of-Work was presented as a business and economic solution to spam. However, as Laurie and Clayton illustrate, Proof-of-Work on its own is not a solution to the problem of spam.[3]

This paper illustrates that Proof-of-Work systems function, in economic terms, when combined with simple price discrimination based on a two-state reputation mechanism. This paper begins by describing Proof-of-Work and providing a review of the derivation of the parameters used by Laurie and Clayton to evaluate Proof-of-Work. The next section provides a brief overview of the current state of the art of deployed anti-spam reputation systems. The core of the paper, an illustration that Proof-of-Work can work with simple price discrimination, is the focus of the next section. A note on Proof-of-

---

[1] Secure Computing, *TrustedSource: The Next-Generation Reputation System*, White Paper, http://www.securecomputing.com/pdf/CT-TS-WP.pdf (accessed October 15, 2007).

[2] In the following paper "proof-of-work" is referred to as "proof-of-effort." Cynthia Dwork, Andrew Golberg, and Moni Naor, "On Memory-Bound Functions for Fighting Spam," in *Advances in Cryptology – CRYPTO 2003*, Dan Boneh, ed., (New York: Springer-Verlag, 2003): 426-44.

[3] Ben Laurie and Richard Clayton, "'Proof-of-Work' Proves Not to Work," May 3, 2004, http://www.cl.cam.ac.uk/~rnc1/proofwork.pdf (accessed October 15, 2007).

Work as currency and the effect of the inclusion of price discrimination follows. Finally, this paper closes by asking why theory and observed reality diverge and discussing how future research will be used to answer that question.

II. DEFINING PROOF-OF-WORK

The core enabling factor of spam is that spam is cheap to send. Proof-of-Work is designed to remove the profit from spam.

Proof-of-Work comprises a set of proposals. Different proposals require that email senders make fungible payments, perform a resource-intensive computation, perform a series of memory operations, or post a bond before sending each message.[4] This section describes the initial Proof-of-Work proposal and details different extensions and analyses of that proposal.

In 1992, the first computational technique for combating spam was presented by Cynthia Dwork and Moni Naor. The fundamental intellectual contribution of their approach was to require that an email sender compute some moderately hard, but not intractable, function of the message in order to initiate transmission. Initiating a transmission means gaining access to resources: the network for transmission, the user's storage in an inbox, and the user's attention span if the transmission is accepted.[5] The essence of Proof-of-Work is: "if you want to send me a message, then you must prove your email is worth receiving by spending some resource of your own."

The current popular Proof-of-Work system is the hashcash system. Hashcash was derived from MicroMint and PayWord.[6] Hashcash is implemented by requiring a sender to determine a hash collision.[7]

---

[4] *See* Cynthia Dwork and Moni Naor, "Pricing via Processing or Combatting [sic] Junk Mail," in *Advances in Cryptology – CRYPTO '92*, ed. Ernest F. Brickell, (New York: Springer-Verlag, 1993): 139-47; Balachander Krishnamurthy and Ed Blackmond, "SHRED: Spam Harassment Reduction via Economic Disincentives," AT&T Labs-Research, http://www.research.att.com/~bala/papers/shred-ietf56-talk.pdf (accessed October 15, 2007).

[5] Krishnamurthy and Blackmond, "SHRED: Spam Harassment."

[6] *See* Adam Back, "Hashcash – A Denial of Service Counter-Measure," August 1, 2002, hashcash.org, 2002, http://www.hashcash.org/papers/hashcash.pdf (accessed October 15, 2007); Ronald L. Rivest and Adi Shamir, "PayWord and MicroMint: Two Simple Micropayment Schemes," April 27, 2001, http://theory.csail.mit.edu/~rivest/RivestShamir-mpay.pdf (accessed October 15, 2007).

[7] Back, "Hashcash."

A hash function is a one-way compression. Imagine a simple function that took all the personal data from a database and reduced it to a set of birthdays. The set of birthdays would be a compression of the data. The function is one way because there would be no way to produce the original data set from the information available in the list of birthdays.

Collisions occur when two different inputs to the same function result in the same output. For example, if the hash functions were "what is your birthday," the "May 6" would be a collision of this birthday hash of the author and Tony Blair.[8] This is because both have birth dates of May 6. Any person can be subject to the "birthday hash." Searching by birth date on the web obviously makes this easy. However, imagine if finding such a collision required asking individuals their date of birth until you obtained two people with the same birthday. Odds are you would have to ask a few more than fifty people. After finding that collision, the person who verifies would only have to inquire with the two people you had found. Thus, the person creating this birthday collision would have to do twenty-five times as much work as the person verifying it.

Similarly, finding collision is difficult with cryptographic hash functions, as many calculations (as opposed to personal queries) are required for one collision. Yet the collisions are cheap to confirm as the recipient needs only to make two calculations. These mechanisms can be used to throttle systematic abuse of un-metered Internet resources, such as email and anonymous re-mailers, in which the sender can be required to compute some processing intensive function, which can be used as a Proof-of-Work.

Of course, the time investment in any processing-intensive Proof-of-Work system depends upon the specific platform. Work that might take twenty seconds on a Pentium IV could take several minutes or more on a Pentium II, and could be infeasible on a smart phone. To address this problem, a Proof-of-Work pricing function based on accessing large amounts of random access memory as opposed to raw processing power was originally proposed by Cynthia Dwork, Andrew Goldberg, and Moni Naor.[9] Later published work identified additional memory-bound mechanisms.[10] Since memory speeds vary much less

---

[8] Cryptographic hash functions take any digital input and generate a non-predictable repeatable output.

[9] Dwork, Golberg, and Naor, "On Memory-Bound Functions."

[10] Martin Abadi and others, eds., "Moderately Hard, Memory-Bound Functions," *ACM Transaction in Internet Technology* 5 (2005): 299–327.

across machines than CPU speeds, memory-bound functions should be more equitable than CPU-bound functions. While processing speeds can vary by orders of magnitude, Dwork, Goldberg and Naor claim a factor of four between the fastest and the slowest memory operations across different platforms. The current Microsoft implementation, Penny Black, is designed to be agnostic about the form of work and requires only some form of work.[11]

Processing cost was the basis of the original model and is the one most examined. This paper concerns itself with the costs of performing some moderately expensive computation as Proof-of-Work, building upon the parameters in "Proof-of-Work Proves Not to Work."[12] The objections to Proof-of-Work before were primarily observations about the high variance not only in the wealth of senders, but also in the processing ability of devices.[13] (Issues of power consumption are another criticism. The dynamic system discussed in future work will address this issue.) Yet, the combination of a reputation mechanism and Proof-of-Work proposed in this paper would work with any of the proposed Proof-of-Work systems.

III. WHAT IS REQUIRED FOR PROOF-OF-WORK TO WORK

Proof-of-Work as a concept appears powerful enough to solve the spam problem by changing the underlying economics of spam. Yet, Laurie and Clayton show that it is not possible to discourage spammers by means of a Proof-of-Work system without having an unacceptable impact on legitimate senders of email.[14] Obviously, simply altering the parameters used in their model would resolve the conflict between spammers and legitimate users.[15] However, such a trivial argument would be neither productive nor engaging. The numbers presented by Laurie and Clayton identify a critical issue, the shift in the production frontier, which must be resolved for Proof-of-Work to be feasible. Therefore their parameters are used to address the feasibility of Proof-of-Work systems. Note that the general models can be used with different parameters.

---

[11] Microsoft Research, "The Penny Black Project," http://research.microsoft.com/research/sv/pennyblack/index.asp (accessed October 15, 2007).

[12] Laurie and Clayton, "'Proof-of-Work' Proves Not to Work."

[13] Ibid., 2.

[14] Ibid., 1.

[15] Ibid., 9.

The following paragraphs review and discuss the parameters calculated in "Proof-of-Work Proves Not to Work." By illustrating that Proof-of-Work can work under those parameters, the specific case is solved. This work illustrates that Proof-of-Work solutions can work, if augmented by a simple reputation mechanism.

To begin a review of the previously determined parameters, recall Radicati's estimation that as of November 2003, on average, 5.7×10^{10} emails were sent and received per day by 5.13×10^{8} email users on the Internet using 9.02×10^{8} email accounts.[16] Brightmail's estimation is that 56% of all emails are spam.[17] Using the Internet Domain Survey's estimation[18] that there were at that time a total of 2.3×10^{8} hosts, Laurie and Clayton concluded that there are 3.2×10^{10} spam and 2.5×10^{10} legitimate emails.[19] This assumes that each machine would send an average of 125 emails per day.[20] From their examination in the UK, Laurie and Clayton further assumed that the proportion of legitimate, non-list[21] emails being sent by each machine is about 60%, thus a final average of about seventy-five legitimate non-list emails being sent is determined.[22] These estimates are accepted.

Using cost estimates of processing power, Laurie and Clayton estimated that the resulting price is $1.75 per machine per day for email operations. Considering spammers used to charge as much as 0.1 cents per email, one spammer must send at least 1750 emails per day to cover his cost.[23] Therefore, the Proof-of-Work calculation time must be 50+ seconds to remove the profit from spam at this price.[24]

---

[16] Radicati Group, Inc., "Market Numbers Summary Update, Q4 2003," news release, November 5, 2003, http://www.radicati.com/uploaded_files/news/Q4-2003_PressRelease.pdf (accessed October 15, 2007).

[17] Brightmail Inc., "Spam Percentages and Spam Categories," 2004, http://www.nospam-pl.net/pub/brightmail.com/spamstats_March2004.html (accessed October 15, 2007).

[18] Internet Systems Consortium, "Internet Domain Survey, January 2004," http://www.isc.org/index.pl?/ops/ds/reports/2004-01/ (accessed October 15, 2007).

[19] Laurie and Clayton, "'Proof-of-Work' Proves Not to Work," 3.

[20] Ibid.

[21] A "list" email is one that is sent to a large number of subscribers, for example SSRN.com email "journals."

[22] Laurie and Clayton, "'Proof-of-Work' Proves Not to Work," 4.

[23] Ibid., 5.

[24] Ibid.

At this point, the critical difference between spam and legitimate email must be addressed. Spammers and legitimate senders of email have different production frontiers. Senders of legitimate email purchase equipment and services on a free and open market. Spammers use botnets which steal electronic services by subverting end user machines. This means attackers break into end user machines and add software that allows them to easily control the subverted machines. In the same manner as professional network administrative software, attackers can easily send one command to many machines. The individual subverted end user machines are called "zombies."

The difference of production frontiers means that spammers and legitimate senders of email have different costs. Laurie and Clayton first estimated that 1.1 million machines might be owned by spammers.[25] The result is a pool of a million machines that could send 32,000 spam emails each per day.[26] Using these numbers, a situation in which only 1% of email is spam means a Proof-of-Work calculation time must increase to at least 346 seconds.[27]

Thus in economic terms, the availability of zombie machines shifts the production frontier for spammers. Spammers have a far lower cost of email production than legitimate users. My proposal uses a two-state reputation mechanism to addresses this difference in cost. In fact, if the difference in the production frontier were on an order of magnitude of a ten or twenty time decrease in cost, the reputation-enhanced Proof-of-Work system described here would still work.

Finally, Laurie and Clayton examined logging data from the large UK ISP. They found that although 93.5% of machines sent less than seventy-five emails per day, a Proof-of-Work mechanism would prevent legitimate activity by the 13% of users who send the most email.[28] In addition, because spammers may select fast machines while legitimate senders are using relatively slow machines, the impact on legitimate email senders could imaginably be worse.

---

[25] Laurie and Clayton, "'Proof-of-Work' Proves Not to Work," 6.

[26] Ibid.

[27] Ibid.

[28] Ibid.

IV. CURRENT ANTI-SPAM REPUTATION MECHANISMS

Proof-of-Work has not been widely adopted as an anti-spam mechanism. Microsoft is endeavoring to change this with the introduction of Penny Black.[29] Currently the anti-spam market is dominated by subscriber services dedicated to blocking or filtering spam. These services include AppRiver, Brightmail, and CipherTrust. This section describes the reputation element of these various anti-spam entities.

In general, the reputation systems for spam are designed to track the history of a sender of email. Different mechanisms are used to track and rate sender behavior over time. Behavior is classified in these systems as good (i.e., sending legitimate email) or bad (i.e., sending spam or malicious mail). Malicious email includes phishing attacks and mail containing malicious code, such as a virus or a worm. Reputation systems may also create profiles for the identification of known historical behavior. For example, a previously trusted account sending out malicious mail may indicate a user who is trustworthy in moral terms (not a spammer), but has been subverted and can no longer be trusted.

The first generation reputation systems used simple blacklists and whitelists. The Real Time Black Hole list is the best known of these simple blacklists. Blacklists contain the IP addresses of known spammers and virus senders and whitelists contain the IP addresses of senders known to be legitimate.[30] Obviously, the first generation of reputation systems had significant room for improvement. For example, a sender's reputation could be affected by the behavior of all senders with whom the sender shared network resources, or a sender's reputation could be affected by malicious code that was sent out with falsified origin or "sender" fields.[31]

Later reputation systems included dynamically updated lists which allowed reputation systems to adjust to rapidly changing conditions and, arguably more importantly, included automatic updates which mitigated the administrative burden of fighting spam. Increasing

---

[29] Microsoft Research, "The Penny Black Project."

[30] Viput Ved Prakash and Adam O'Donnell, "Fighting Spam with Reputation Systems," *ACM Queue*, no. 9 (2005), http://acmqueue.com/modules.php?name=Content&pa=showpage&pid=346 (accessed October 15, 2007).

[31] O'Reilly & Associates, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, ed. Andy Oram (California: O'Reilly, 2001).

storage and processing power enabled more granular message scoring. Blacklists were replaced with per-email numerical scores indicating probabilistic weighing of the likelihood of spam.[32] Modern anti-spam mechanisms are difficult to evaluate in detail because the mechanisms for weighing and storing reputations are as much business intelligence as they are art or science.

Researchers have created sets of requirements for reputation systems and argue that an effective reputation system must be dynamic, comprehensive and precise.[33] Anti-spam reputation systems must be based on actual enterprise mail traffic in order to keep the spammers from gaining any advantage.[34] Today, the latest reputation systems take a persistence testing approach to reputation scoring. Some systems also evaluate the social network of the sender to determine reputation scores. Both CipherTrust and Gmail have significant information about the social network of recipients who subscribe to their services.

Despite the existing commercial differentiation of systems, there is a common core to the anti-spam reputation systems. Most of the existing reputation mechanisms use the average of past feedback reports to assess the reputation of one agent.[35] Agents may be as broad as a domain, based on IP address, or as narrow as a single email address. Different reputation system providers have differing characteristics, and therefore, differing cost functions and error rates. The error rates published by commercial providers may be goals as much as they are historical measurements.

The critical observations for this work is that reputation systems, which function on a per-email, per-address or per-domain basis, already exist. Complex rating mechanisms as well as historical reputation mechanisms are currently used in commercial anti-spam technology. The mechanism proposed here is not unduly complex in comparison with current anti-spam products.

---

[32] Prakash and O'Donnell, "Fighting Spam."

[33] R. Jurca and B. Faltings, "Reputation-based Pricing of P2P Service," in *Proceeding of the 2005 ACM SIGCOMM Workshop on Economics of Peer-to-Peer Systems* (New York: ACM Press, 2005): 144-49.

[34] Prakash and O'Donnell, "Fighting Spam."

[35] Ibid.

V. PROOF-OF-WORK AUGMENTED WITH PRICE DISCRIMINATION

This paper proposes a simple model combining a reputation mechanism and Proof-of-Work scheme, which enables price discrimination. The system is a step function: each email sent has either a high or a low Proof-of-Work requirement. Effectively, this paper proposes placing a corresponding low cost on known reliable parties, with a high cost on those identified as spammers. Based on an assumption that one zombie machine can be detected during several minutes, the high Proof-of-Work cost requirement is implemented after the first detected spam and held for a set duration. Emails are rated based on a per-email or per-source basis.

Newcomers are overwhelmingly malevolent in the world of SMTP[36] servers. The research done by CipherTrust identified approximately 50 million IP addresses which send approximately 70% of all email on a daily or near daily basis.[37] The other 30% comes from IP addresses that have not been previously encountered. More than 95% of that 30% of emails from new or unknown IP addresses is malicious.[38] In other words, an IP address that is encountered for the first time is approximately 95% likely be a zombie machine.

One way to detect a zombie machine quickly is to assume that each new entrant is malicious until proven otherwise, as is common in reputation mechanisms.[39] Therefore, this paper proposes placing a high cost on new entrants as well as identified spammers.

My model can be described as follows. When a newcomer arrives, he or she will bear a high Proof-of-Work cost, *H*. The cost of any new or previously malicious new email source will not be fixed at the initial high cost forever. Newcomers can overcome the initial bad reputation and associated higher cost by performing the Proof-of-Work as required and sending only legitimate emails. After bearing this cost when sending the first several emails, his or her reputation will improve if there is no spam detected in his or her sent email. As a result, the Proof-of-Work cost drops immediately to a much lower

---

[36] SMTP stands for Simple Mail Transfer Protocol. SMTP is used for the vast majority of email.

[37] Secure Computing, "Persistence Testing – Guilty Until Proven Innocent," *TrustedSource: The Next-Generation*, 5.

[38] Ibid.

[39] Eric J. Friedman and Paul Resnick, "The Social Cost of Cheap Pseudonyms," *Journal of Economics & Management Strategy* 10 (2001): 173–99.

level, *L*. However, once one sent email is indicated to be spam, the per-email Proof-of-Work cost to this sender will immediately increase back up to *H*. After that, at any time a spam message is detected, regardless of the nature of the subsequent emails, the following emails will bear the high cost *H* as a punishment for some duration after the last detected spam. The initial high Proof-of-Work cost should be high enough to prevent any new entrant from being a profitable zombie.

In game theory terms, the proposed model can be seen as a tit-for-tat model with forgiveness. Defection, in this case of sending spam, results in immediate punishment in the form of increased work. If the participant then behaves well for the following emails, there is "forgiveness." Therefore, a user who is wrongfully identified as a spammer will not pay an indefinite price. Of course, in this simple model the existence of blacklists is not addressed. Clearly, once a participant has been repeatedly identified as a spambot, no email would be accepted.

To examine this proposal, a simulation was developed to examine the average Proof-of-Work cost to end users who are ill or well behaved. In this simulation each was an event that was associated with a probability. The probability decided the cost of each email using two possible vaules: the cost of spam was *H* and the cost of legitimate email was *L*. The various values assigned to the duration of punishment were also considered. Email that was rejected for inadequate Proof-of-Work was bounced to the sender.

The rates of error in spam detection were also considered in the simulation. There were two error rates: one reflected the probability of the false identification of legitimate email as spam; the other reflected the incorrect identification of malicious email as legitimate. According to vendor reports, software exists that can detect spam with a level of accuracy which ranges between 92%-99%.[40] Again, based on vendor claims, the probability a legitimate email may be mistakenly indicated to be spam is approximately 1%. Vendors' tolerance of error types varies. Some vendors never throw out legitimate email but detect less spam. Others detect more spam but lose the occasional email.

Therefore, in this simulation, a single spam email was detected with 99% accuracy while a legitimate email was mischaracterized at a rate of 1%. The resulting expected cost for a spammer to send each spam was approximately 349 seconds, which is close to Laurie and

---

[40] Ian "Gizmo" Richards, "How to Reduce Spam," *Support Alert Newsletter*, November 2006, http://www.techsupportalert.com/how_to_reduce_spam.htm (accessed October 15, 2007).

Clayton's estimate for the time required to discourage spammers.[41] For legitimate users, with 1% detection error, the expected cost of sending each email was around fifty-two seconds. Again, this meets the requirement that end users who send legitimate email are in fact able to do so.

These results suggest that this Proof-of-Work model combined with a step-wise reputation mechanism can work to discourage spammers without overloading high-volume legitimate users.

VI. THE NATURE OF PROOF-OF-WORK

There is a fundamental distinction between Proof-of-Work as generally described and Proof-of-Work with the proposed reputation system. Proof-of-Work as initially described was a token currency. Recall that money is a mechanism of exchange, a store of value and a standard of value. With token money, the value is inherent to the mechanisms of exchange and an exchange of a token is an exchange of value. Token money is either inherently valuable or represents value so that a token is not a function of the party exchanging it. The dollar is an example of a token currency.

In contrast, notational money is exchanged based on notations in a record-keeping system. A credit card charge is an example of notational currency. Notational exchanges are not completed until verified by the record-keeping party. In this case, Proof-of-Work exchanges require the reputation-tracking party to verify that the payment is adequate and thus valid. Penny Black, an on-going research project by Microsoft, is a completely notational implementation of Proof-of-Work.[42] Each user has an account, and each email recipient decreases or credits that account. The model allows those who fight spam to select costs based on the level of granularity that is most effective. End users can keep history-based records and bounce email without Proof-of-Work. ISPs could also keep such records so that the individual user history is not an issue when sending mail.[43] Penny Black is a traditional notational

---

[41] Laurie and Clayton, "'Proof-of-Work' Proves Not to Work," 6.

[42] Microsoft Research, "The Penny Black Project."

[43] Martin Abadi, and others, eds., "Bankable Postage for Network Services," *Advances in Computing Science – ASIAN 2003* (New York: Springer-Verlag, 2003): 72–90.

instantiation that associates each email with exactly one Proof-of-Work account.[44]

Note that there is no requirement that the record-keeper and the parties to the exchange are indeed distinct. There must be a notational clarification for the Proof-of-Work to be accepted. An example of where reputation-based Proof-of-Work would have a single party evaluating and pricing might be in a distributed denial-of-service (DDoS) attack, which is an attack to make a computer resource unavailable to users by multiple compromised systems. Those parties that have some history of transaction – either as identified by a DDoS "cookie" or through another record of interaction – can pay the lower cost. Those parties with no history will be required to pay the greater Proof-of-Work. Indeed, an initial challenge can be easily modeled under this system. The first response requires some "payment," meaning the duration is just one and the payment is equivalent to the challenge. Notice that the duration "one" is indicated above as being inadequate for the email case.

In contrast, a message from a whitelisted individual is the extreme case where the low cost is zero, e.g., $L=0$. The simulations illustrate that the case where the low cost is zero and the high cost is high (>400 seconds) would indeed function within the parameters given. In this case, users either pay a very high Proof-of-Work cost to send email or are members of a whitelist. Recall individuals could each maintain their own whitelists. Those who would initiate conversations would then either pay a premium or obtain an introduction.

Proof-of-Work can work. However, Proof-of-Work requires some notational elements to function in a world where it is impossible to distinguish prima facie between the legitimate and criminal markets.

VII. FUTURE RESEARCH

The modeling in this paper illustrates that Proof-of-Work would work if the work factor were high for spammers and low for known users. Multiple mechanisms that are not traditionally considered Proof-of-Work can fit under the Proof-of-Work rubric. Examples of this include challenge and response mechanisms that require anyone who is not part of the history of the recipient to respond to an email or perform some work for the email to be received. Yet, these mechanisms have not proven to reduce global spam.

---

[44] Microsoft Research, "The Penny Black Project."

The next level of research will be on market dynamics.[45] In the model presented here, and in all other models of Proof-of-Work, there is an assumption that Proof-of-Work is ubiquitous. The assumption of instant, uniform, adoption is common to both computer science and economics and extremely rare in the world on which these sciences are focused. This model suggests the addition of a dynamic element to it so that at any point in time, the number of individuals who adopt a system is a function of the number of users at the previous time.

This paper proposes using the standard equations for a natural dynamic system. However, in these systems, rate of infection, rate of recovery and mortality may be known. Also, due to birth and death, the number of participants changes over time (e.g., equation 3 in footnote 45 will not hold). In Proof-of-Work these are complete unknowns. Thus, the model can be updated based on market adoption and network observations.[46]

Note that the diffusion measure in this dynamic model is correlated with the likelihood of spam detection in the probabilistic model above. This is because some people use Proof-of-Work, then the other people who do not use Proof-of-Work will never detect spam under the definition that spam requires Proof-of-Work.

The dynamic nature of Proof-of-Work provides one possible explanation of why Proof-of-Work is not currently effective: because the expected probability of adoption is arguably well under 60% even considering the ad hoc adoption of Proof-of-Work equivalents.[47] The

---

[45] To be more specific, set number of users of POW to *POWu* and the number of users who do not as *POWn*. At any time, some percentage of users will reject POW, *POWr*, and others will adopt POW, *POWa*.

$$POWu[t+1] = POWu[t] + POWa[t] - POWr[t] \quad \text{(eq.1)}$$

$$POWn[t] = POWn[t] + POWr[t] - POWa[t] \quad \text{(eq.2)}$$

While the total number of users does not change,

$$\text{e.g., } POWn[t+1] + POWu[t+1] = POWn[t] + POWu[t] \quad \text{(eq.3)}$$

[46] *See* Indiana University Network Operations Center, "Network Operations Center Services," March 1, 2001, http://www.indiana.edu/~uits/telecom/noc/ (accessed October 15, 2007).

[47] Individual decisions on accepting or rejecting POW depend on its ubiquity of adoption of POW.

$$POWr[t+1] = -\alpha POWu[t] + \beta POWn[t] \quad \text{(eq.4)}$$

$$POWa[t+1] = \gamma POWu[t] - \delta POWn[t] \quad \text{(eq.5)}$$

modeling of the market dynamics provides one way to test this hypothesis.

One implementation that is based on individual accounts is Penny Black by Microsoft.[48] There are reasons not to adopt Microsoft's Penny Black mechanisms unrelated to network effects or interoperability.[49] Penny Black uses a notational mechanism whereby the participants must have a mutually trusted server (or set of servers) that issues per-email tickets. Email recipients then contact the centralized server again to determine if the ticket is valid. This allows for per-user pricing. However, depending on the implementation, this has the potential to allow Microsoft unprecedented levels of social network information, information on internal corporate communications, and other information from traffic analysis. Of course, Penny Black does not require that an "identity" be linked to an account, only that an email address is linked to an account. While the potential for anonymous accounts is built in, its actual usability and anonymous strength is uncertain. Certainly, no company competing in any market with Microsoft would be interested in providing such information, and end users may be similarly loathe to provide such personal details. The observation of the diffusion of the Microsoft Proof-of-Work mechanism Penny Black will enable, over time, an empirical measure of these constants.

VIII. CONCLUSIONS

Regulatory efforts to stop spam must address a variety of business practices, expectations of customer/merchant relationships, cultural traditions of sales, and enforcement against suspect actors. These efforts must be coordinated at a global level, as even one regulatory spam haven would prevent effective prosecution. Given the difficulty of regulation of far worse crimes, as extreme as extermination of endangered species for profit and the traffic of human beings, spam is unlikely to be prevented by regulation alone. Unlike these heinous crimes, spam is a numbers game. Spam response numbers are relatively low; therefore, the efficacy of spam depends on being able to blast out large amounts at a low price. If Proof-of-Work causes spammers to become more targetable, the high overall cost of spam to the network and society as a whole is greatly reduced. It is the mass

---

[48] Microsoft Research, "The Penny Black Project."

[49] Martin Abadi and others, "Bankable Postage."

nature of spam that damages the network, and it is exactly that characteristic that makes it possible for Proof-of-Work to in fact work.

Proof-of-Work reverses the cost model of email by charging the sender instead of the user. This paper has proposed the combination of Proof-of-Work and a simple reputation mechanism. This paper illustrated that, for legitimate email users, the cost is acceptable; for spammers, the cost is prohibitive. Using multiple simulations, the paper illustrated that Proof-of-Work with a simple reputation mechanism can work over a wide range of values. The result raises as many questions as it answers, and a research agenda is offered as a method of moving forward.

Recall that a uniform Proof-of-Work mechanism will not work because any price high enough to stop malicious email will be so high that it will hinder legitimate users. In fact, the low cost of stolen network goods (e.g., botnets) requires that the cost to a spammer be an order of magnitude higher than the cost to a legitimate user for Proof-of-Work.

This work examines Proof-of-Work as part of a larger anti-spam effort. Current anti-spam vendors use reputation systems as well as per-email spam evaluation mechanisms. These efforts suffer from penalizing new IP addresses and discarding incorrectly identified email. The types of error are difficult to balance. Either new entrants are not allowed to send email, or each new IP address is allowed to send enough email that spam remains profitable. Proof-of-Work can be combined with per-email spam identification and source reputation to create more effective anti-spam technologies.

Proof-of-Work can work, using the economic conditions derived as necessary from previous work. In summary, this article has examined Proof-of-Work as an element of anti-spam technologies as combined with source identification or per-email evaluation. As such, Proof-of-Work could work.