IPv6 and Spam

Peter Kosik, Patrik Ostrihon and Reza Rajabiun*

2009 MIT Spam Conference

Keywords

spam, internet protocol, multilayer filtering, technological diffusion

Abstract

Implementation of Internet Protocol version 6 (IPv6) appears necessary for further growth and development of integrated communications networks. However, this important enhancement to the Internet Protocol has failed to gain widespread acceptance. This paper highlights that the persistence and sophistication of spam reduces incentives to deploy IPv6. The analysis suggests that IPv6 adoption will erode the efficiency of antispam mechanisms that classify communications based on the reputation of its senders and will require an increased emphasis on content filtering.

I. Introduction

The rapid decline in the cost of telecommunication has provided the basis for large scale efficiency enhancements in almost all other sectors of the economy. This is because good communication technologies lower search costs that buyers and sellers must incur to find each other and engage in Pareto efficient transactions. In the past decade the emergence of spam on email systems has demonstrated that when costs of using open communications networks go down, some individuals and groups have strong incentives to abuse their access privileges. While volatile on a daily basis, on bad days the volume of email content defined as undesirable by end users or their filters can reach 90% of total messages trying to enter a network. [1] Spamming is also increasingly common on mobile interfaces that are becoming the primary method for connecting to the Internet.

One consequence of the growth and sophistication of spam has been the development of highly accurate statistical content classifiers that can filter 98-99% of the noise. [2] Advances in this area mitigate the costs that spam poses on the scarce resources of end users of email. Clearly, without these filters end users would have difficulties employing email and other low cost messaging platforms. The high noise to signal ratio nonetheless imposes significant network costs on service providers in terms of hardware, human resources, bandwidth and storage facilities. A recent study by the International Telecommunication Union (ITU) compiles and reviews international evidence on the costs of spam and messaging abuse on end users and network operators. [3]

This paper highlights a more pernicious cost of spam than those outlined in existing studies. It hypothesizes that the capacity of spammers to produce large amounts of undesirable messages conditions technological decisions by service providers that connect end users. If this is the case, existing studies are likely to be significantly underestimating the costs of spam and its implications for further development of the Internet. We explore this hypothesis in the context of the proposed transition to Internet Protocol Version 6 (IPv6). The analysis suggests that mechanisms that aim to classify messages based on the reputation of senders are likely to be less effective under the new protocol. Such mechanisms are nonetheless in widespread use, particularly by upstream providers of network connectivity. [4] Deployment of IPv6 consequently erodes the relative capacity of administrators to defend their networks against spammers. This observation reveals a potentially

* COMDOM Software, Toronto, Canada
Corresponding author:
<reza@comdomsoft.com>

complementary relationship between IPv6 diffusion and increased emphasis on content scanning and classification in multilayer spam filters. [5] This paper explores how risk aversion resulting from the spam problem constrains the prospects of widespread IPv6 diffusion.

As illustrated in Figure 1, issues raised in this paper are of increasing importance because of recent growth in demand for IPv6 connectivity. The trading data further reveal a high degree of volatility in IPv6 traffic, which potentially highlights its experimental status and low levels of liquidity in the open market. In the past year, industry bodies such as the Messaging Anti Abuse Working Group (MAAWG) have also set up related technical sub-committees to study various aspects of this issue. By integrating the economic and technical perspective, this paper aims to provide a basis for further research in this area. [7]

about the reputation of senders. The last section draws inferences from the analysis for the diffusion of IPv6 and optimization of multilayer filtering systems.

II. IPv6 economics

IPv4 represents the first widely employed standard for identifying, routing, and maintaining the integrity of packet-switched internetwork. The design and deployment of IPv4 was conducted within relatively hierarchical public sector organizations, notably the U.S. Department of Defense (DOD). Hence, the environment for this successful bundle of protocols was significantly different than the one facing IPv6 implementation today. IPv6 is a first major attempt by non-state actors, specifically the Internet Engineering Task Force (IETF), to introduce a core standard necessary for further growth and development of the network (RFC 2460, 1998).[8] Given that the increasingly

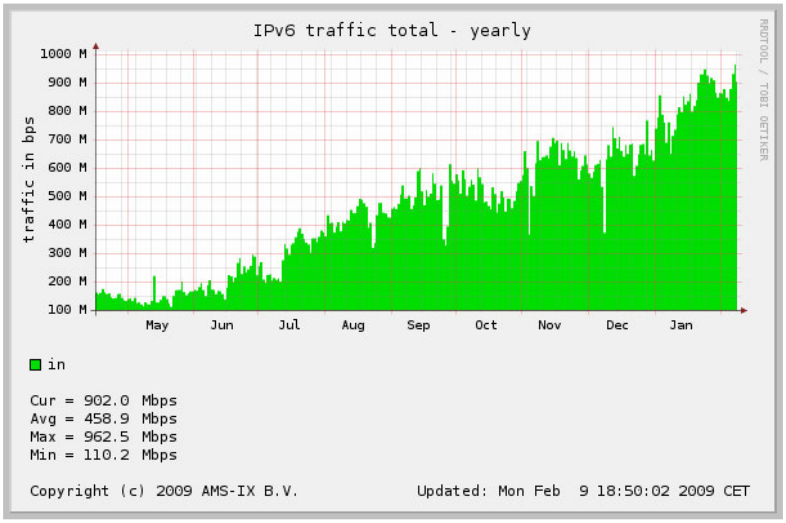

Figure 1 – *Source: www.ams-ix.net/technical/stats/*

The first part of the paper reviews theoretical and empirical work on the emerging IPv6 network. It further highlights how the interdependence of investment decisions by Autonomous Systems (AS) can constrain the diffusion of such enhancements. The second section studies how the technical features of IPv6 are likely to shape the capacity of spammers to search for targets and bypass antispam mechanisms that rely on information

distributed nature of the Internet makes a global mandate impractical, the widespread implementation of the enhancements to the protocol faces significant challenges.

Proposals culminating in the development of IPv6 predate the tit-for-tat battle between spammers and network operators. As such, the motivations for adopting main elements of the bundle are nominally independent of the growth in volume and sophistication of spam

in email and other messaging networks since the early 2000s. As documented in this paper however, the interdependence of decisions by network operators and spammers is likely to influence the IPv6 diffusion path. In order to focus on the implications of the spam problem, the analysis that follows abstracts away from a discussion of administrative, hardware, or infrastructure costs that also shape decisions about a new technology or standard.

Demand for IPv6: Despite its resilience and flexibility, by the late 1990s the 32 bit design of the IPv4 address space started to present a challenge to further expansion of the Internet. The demand for IPv6 can be viewed in terms of the opportunity costs involved in retaining IPv4, particularly in terms of quantity and quality of end to end connectivity across semi-autonomous network operators. An obvious solution to the address exhaustion problem is to create a market that prices the scarcity of IPv4 space, and hope that this stimulates IPv6 adoption. This type of mechanism does not seem to be a practical option however.

connections between an almost unlimited number of individual physical and virtual machines with unique addresses. Other refinements, such as changes to packet header formatting and standardization of IPsec are also part of IPv6. While any one of the technical features included in the bundle may be relevant in the context of the spam problem, in this paper we focus only on the implications of the radical increase in the size of the address space and autoconfiguration functionality of the standard for the relative capacity of spammers and network providers.

The fact that IPv6 was codified by IETF only five years after the privatization of the Internet suggests that the prospects of address space rationing under IPv4 have been apparent for more than a decade. This has been a problem for the sustainable growth of connectivity, particularly in developing countries where mobile devices have radically expanded access in the past years. Diffusion of IPv6 would address quality and quantity of access problems associated with asymmetries in the distribution of address space under the current

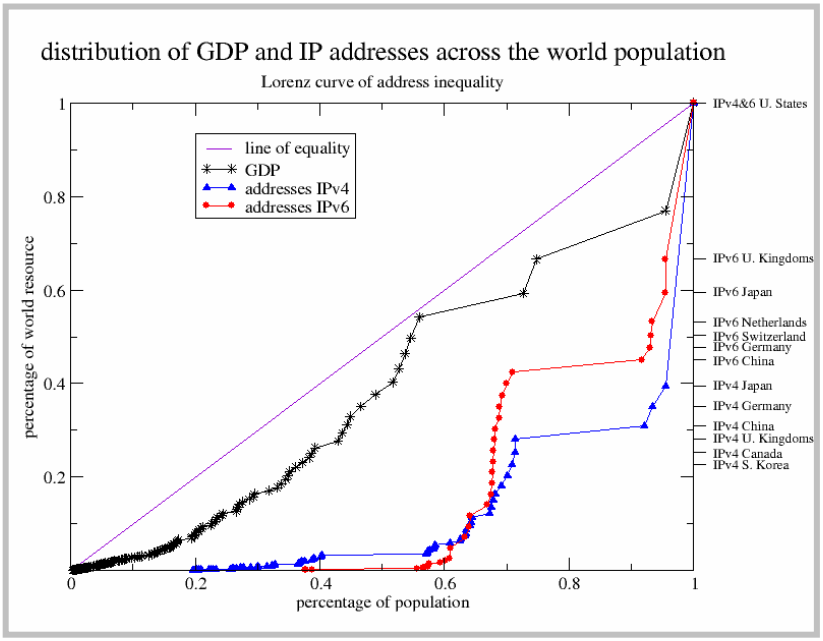

***Figure 2** – Source: www.caida.org/research/policy/geopolitical/bgp2country/ipv6.xml*

The size of the address space under the IPv4 standard provides just over 4 billion unique addresses, chunks of which are reserved for military networks. The use of a 128 bit address in IPv6 allows for the possibility of end to end

standard. [9]

In addition to alleviating the address exhaustion problem, the vast amount of unique addresses can be employed by operators to

provide innovative services that require distributed command and control of a large number of hosts and clients. Furthermore, hosts employing IPv6 may configure themselves using ICMPv6 router discovery messages (stateless address auto-configuration) and communicate with their neighbors in finding efficient routes for the underlying traffic. This makes the standard attractive to network providers because it allows them to offer new services and optimize the use of their infrastructure. For instance, with IPv6 cable modems can be assigned unique addresses that make it easier to manage and deliver end user connectivity across a wide range of platforms. [10] The example of the U.S. Department of Defense is again instructive. While this organization has chunks of reserved IPv4 space, it has been one of the leaders in deploying IPv6. The potential for delivering new services with IPv6, ranging from drone controls to mobile messaging in rural areas, is one driver for its diffusion. [11]

Persistence of IPv4 also has notable costs in terms of the degree of integration of the network. Specifically, Network Address Translation (NAT) has been widely used in developing countries to accommodate growing demand for access. Since NATs obfuscate underlying network information and inherently isolate an autonomous network, this solution to IP address scarcity significantly limits the quality of end to end connectivity. The resulting fragmentation of national networks and reliance on walled gardens of private networks further justifies concerns motivating IPv6. Importantly, IPv6 to IPv4 address masquerading can be employed by operators to solve IP address scarcity within a particular network. The AS could then communicate with the outside world using IPv4. The so-called "IPv6 lite" approach to implementation is increasingly common, but it is not likely to stimulate widespread diffusion of IPv6 across network providers. [12]

Supply of IPv6: Trading data illustrated in Figure 1 revealed significant improvements in the liquidity of the global market for IPv6 traffic since the middle of 2008. A wide range of public and private sector projects have contributed to the development of the emerging IPv6 network. Figure 3 maps the geographical location of ASs deploying IPV6 based Dolphin network topology discovery system. [13] Overall, as of December 2008, the system had identified 631 ASs, 15769 routers that provided IPv6 connectivity.

Despite the presence of a basic IPv6 network and increasingly liquid trading, the overall level of penetration of the technology remains very low. A 2008 study found for example that

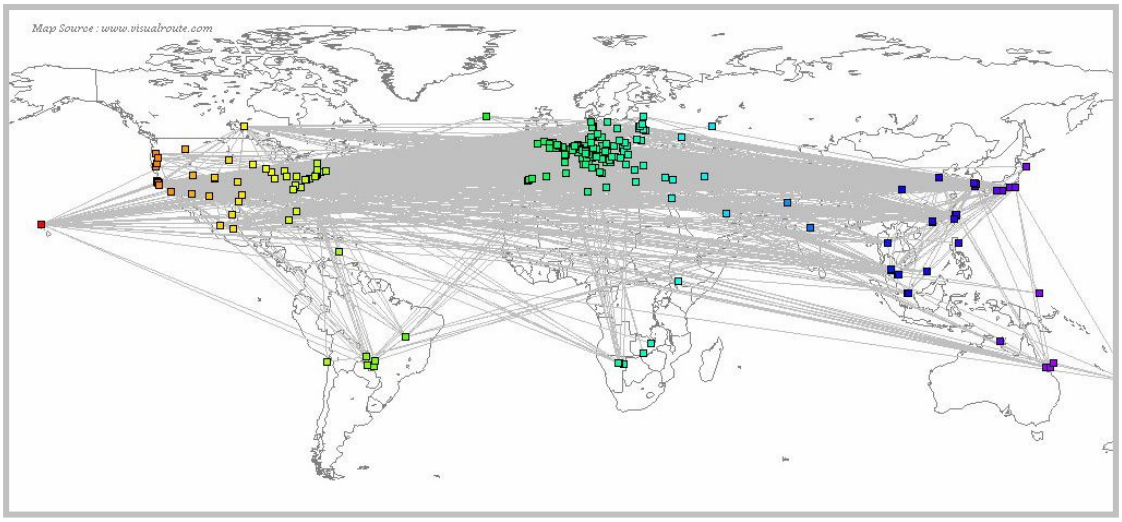

***Figure 3** – Source: http://ipv6.nlsde.buaa.edu.cn/*

only 0.238% of end users had IPv6 connectivity. [14] Highest levels of penetration were present in Russia (0.76%) and Eastern Europe, and the lowest in Japan (0.15%) and Asia. This suggests significant impediments to the adoption of IPv6 connectivity and the difficulties in generating demand for the standard using traditional policy tools, notably subsidies and partial mandates. Minimal penetration rates remain even in countries that are likely to gain more from the diffusion of the new standard and have tried to stimulate its adoption.

In this context, it seems reasonable to look for common technical and economic factors that can influence decisions about the standard. The obvious candidates that limit IPv6 diffusion are the fixed costs of upgrading equipment and administrative capacity to operate under the new protocol. Another reason may be low end user demand for IPv6 connectivity, which would in turn constrain incentives of network operators to incur fixed costs associated with upgrades. In particular, the most common operating system – Windows XP with approximately 64% market share at the end of 2008 - does not support IPv6 by default. [15] Windows Vista has "unified" communication capabilities supporting both standards by default, but it has not been implemented widely yet. If their customers are not ready to exploit the new standard, it might be very inefficient for network providers to invest in IPv6 infrastructure today. They might wait for demand to develop, or delay its implementation in the expectation that technological change will reduce the price of IPv6 infrastructure in the future.

In addition to the usual supply and demand factors, resistance to the adoption of the new technology may be viewed in terms of strategic incentives by incumbent service providers to limit potential entrants to their markets. Incumbent carriers are less likely to have IPv4 address space problems than emerging operators, and hence might have strategic reasons to retain the old standard as an indirect instrument for limiting competition. While this paper does not explore this issue further, it is relevant to point out competition considerations, as well as many other issues not discussed here, since they also inform decisions about the adoption of IPv6.

IPv6 Diffusion: The idea of diffusion has been studied in a wide range of natural and social sciences, providing a wide range of potentially relevant models for the analysis of IPv6. Traditional models assume that initially only a small sub-group of innovators implements a new technology. They also assume that with time the rest of the consumers learn to adopt a particular efficiency enhancing technology. This deterministic perspective generates the well-known diffusion S-curve. Elmore et al. (2008) apply this model to the analysis of IPv6 diffusion. [16] Pointing out that IPv6 is unlikely to be adopted in a timely manner, they argue that a combination of subsidies, regulatory mandates, and technology bundling will be necessary to stimulate its adoption.

More generally, the notion of positive network externalities is often used to explain underinvestment, for instance in education or public healthcare. If social benefits of a particular decision, such as going to school or getting a flu shot, are larger than its private benefits, decentralized decisions can result in inefficiently low levels of private investment in that area. This is because the value of making the "right" choice for individuals goes up as the rest of the population also makes the same choice. In the IPv6 case, this framework implies that when overall adoption rates are low, private benefits of investing in the new standard may appear too low to individual network operators to justify the fixed costs of upgrading infrastructure and administrative skills.

More generally, the IPv6 problem can be usefully characterized in the language of game theory. Specifically, in the presence of positive network externalities, decisions to adopt/not adopt IPv6 by individual network operators and enterprises are likely to reinforce each other. This would mean that the decision to invest in IPv6 by some ASs would increase the probability of others doing the same. Analogously, a strategy of delay in the relevant investments by some entities

reinforces others to wait also. This type of reasoning generates the well known problem of dissonance between Nash and Pareto optimal equilibria in non-cooperative games. In the case of IPv6, this approach suggests that an inefficiently low penetration level is likely to represent a stable equilibrium.

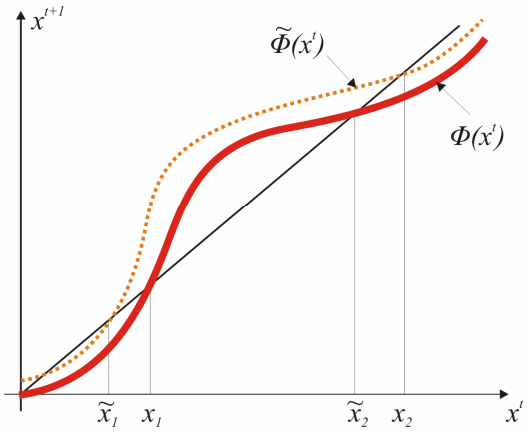

Figure 4 – *Source: Jackson and Yariv (2007)*

Jackson and Yariv (2007) provide a general model of diffusion of binary decisions in network games. [17] They show that when payoffs to choices of this type exhibit strategic complementarities, a sup-optimal Bayesian-Nash type equilibrium exists, and that convergence of behavior is monotone either upward or downward. Following the standard language in network sciences, let x^t represent the (link-weighted) fraction of networks that have adopted IPv6 at time t. For the analysis of different possible decision processes by network operators, also let d_i represent the degree distribution of network i, or the number of links between a particular network and its neighbors. Let x represent the probability that a particular network will adopt IPv6. The payoffs from adopting IPv6 v then depend on expectations of each operator about those by her neighbors. An AS will adopt the new technology if the cost of adoption c:

$$c_i \leq v(d_i, x)$$

Different processes for expectation formation are plausible and will determine which equilibrium (high/low diffusion) is achieved, its stability, and tipping points for transition between the two states of the world. Consider the following possibilities:

- When an AS only cares about the average play of her neighbors, and network structure does not matter:
  $v(d,x) = u(x)$
- When a network's individual payoff to IPv6 is a function of expected number of neighbors adopting IPv6:
  $v(d,x) = u(dx)$
- When $v(d,x)$ is a step function, so that the decision to adopt IPv6 takes place only if x lies above a specific subjective threshold.

The first case represents the traditional models of diffusion which do not take account of network structure. The second perspective suggests that the diffusion process is likely to gain momentum only after large backbone network providers, with a large number of linkages with outsiders (large d's), have made the first move. This would increase the private benefits of adopting the standard by enterprises and other downstream networks, and potentially motivate them to switch. The last case represents the problem as it is likely to look to an upstream operator that must incur the bulk of fixed investments. A stepwise relationship in the value of the technology suggests that a shift between low and high states of diffusion is possible with small perturbations to the system. An inefficiently low level of penetration under these conditions can be highly persistent and robust to impositions of instruments such as taxes on IPv4 or subsidies for IPv6. The invariance of the diffusion rates to recent bootstrapping efforts, namely setting up IPv6 hubs, administrative mandates or subsidies, highlights this possibility. [18] Decisions by network operators with the largest number of linkages are likely to be central to the widespread diffusion of the enhancement.

While low penetration rates of IPv6 make empirical verification difficult, Xiao et al. (2009) provide some first insights that support this hypothesis.[20] They found that the degree distribution of IPv6 ASs follows a power law distribution, as other scale free networks. They estimate the exponent of the

degree distribution - which measures the uniformity of degree distributions in the network - to be around 1.2. This estimate is much lower than those found in the IPv4 AS topology (around 2.2) and many other social and digital networks. The implication is much less uniformity in the distribution of linkages across ASs in the current IPv6 network, which they explain using a theoretical model that emphasizes probability of preferential attachment and edge rewiring under IPv6.

opportunity costs for the growth of digital communications. This is particularly the case in developing countries where financial and technical expertise are scarce relative to the industrialized world.

Recent research suggests that the growth in spam volumes is not only costly to end users and network operators, but may not be very economical for spammers either. For instance, Kanich et al. (2008) infiltrate the command and control infrastructure of a widely used

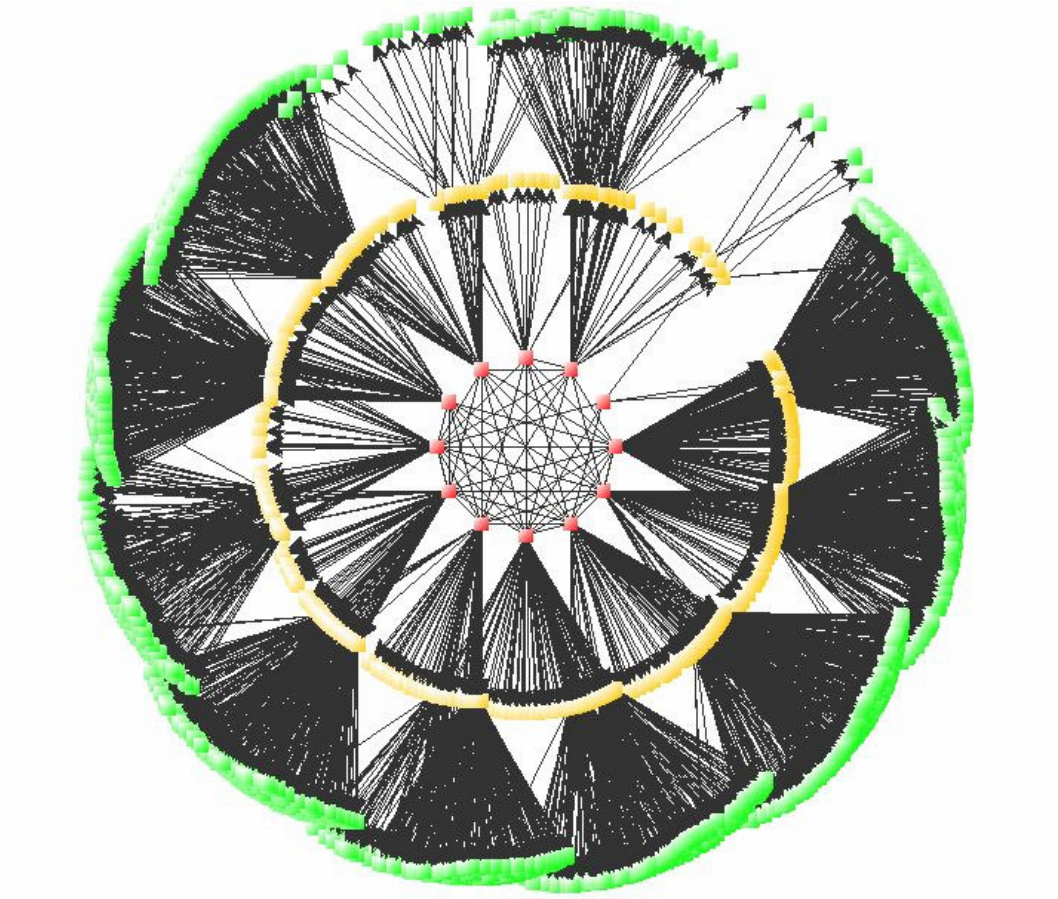

Figure 5 *– Source: http://ipv6.nlsde.buaa.edu.cn/*

III. Spam control in transition to IPv6

While filters take out most of the spam messages before they get to end users, costs of spam in terms of network infrastructure and human resources by network providers remain substantial. Moreover, by draining resources from more productive projects that can expand the quality and quantity of access, spam has

spam robot and find that the conversion rate for their spam in terms of either sale of advertised products or downloads of malware Trojan was much lower than previously believed. [21] Their experimental campaigns suggest that the value of each target to a spammer is now extremely low, yet the high ratio of noise to signal that emerged in the early to mid-2000s continues to persist.

To capture this puzzle, consider this issue in terms of a basic prisoners' dilemma. [22, 23] Both attackers and defenders know that in aggregate they would be better off if they deescalated their battles. If they produced less spam and further personalized their advertisement, spammers may be more successful in avoiding detection by self-learning spam filters and increase their expected conversion rates. However, the fact that the costs of sending large amounts of messages are so low motivates advertisers to build mailing lists that are larger than optimal and invest too little in personalization and targeting of their messages. Some degree of cooperation between advertisers and network operators can help address the problem to some degree, but will not change the underlying economic incentives that generate the system failure.

O'Donnell (2008) argues that the central factor in determining the strategy of attackers is the value of rents they can extract from users over time. [24] This factor depends on both the preferences of end users in a particular type of messaging networks (e.g. email, MySpace, Facebook, Twitter), as well as the quality of malware produced for a particular system vulnerability (e.g. Windows versus Mac). In the short term, IPv6 represents a new opportunity for spammers to search for network vulnerabilities and potential targets. Given that the low response rates present today do not constrain the production of spam, it seems prudent to assume that in the longer term they will have strong incentive to exploit the protocol. As well as exploring how they can subvert specific features of IPv6, developers of spam can also use the transition process between the two standards to construct novel angles of attack.

In the longer term, two basic features of IPv6 are likely to be instrumental in shaping spammers' technologies and strategies: a) Its extremely large address space, and b) Its autoconfiguration functionality. The rest of this section explores linkages between IPv6 and spam by focusing on the implications of these features of the bundle. Specifically, we study how they are likely to shape decisions by end users, different types of network operators, and spammers.

End users: Decisions about the version of the Internet Protocol are not usually made by end users. At this level, the widespread diffusion of the new standard requires that IPv6 support is enabled by choice or default. Since most end users do not actively pick which protocol to deploy to identify them and route their packets, the adoption of desktop operating systems and applications that automate IPv6 connectivity is a necessary condition for its widespread diffusion. At some point, the penetration level of OSs that automatically support both standards (Windows Vista in particular) will reach some threshold level where it becomes profitable for network operators to invest in IPv6 infrastructure. If the new protocol results in new and costly vulnerabilities for end users, it is possible that they could disable its support and reduce their demand for IPv6 connectivity from their providers.

While necessary, the transition to IPv6 capable OSs at the end user level may not be sufficient to stimulate investment by network providers or diffuse the bundle. Even if a large proportion of end users had fully automated IPv6 capabilities installed, financial or security considerations can reduce investment by network providers in connectivity via the new protocol.

Downstream operators: Enterprise network managers often acquire bundled messaging applications. Since they typically provide the first link between end users and the Internet, their reaction to IPv6 security vulnerabilities might be relevant to the diffusion of the protocol from the edges of the network to its core infrastructure. If downstream servers are not open to IPv6 traffic because of spam or DDoS concerns, end user level capabilities are likely to be less of a relevant factor. The Microsoft Exchange 2007 Antispam and Antivirus Functionality documentation helps illustrate the implications of the spam problem at this level of the network. The vendor:

"…strongly recommend against configuring Receive connectors to accept anonymous connections from unknown IPv6 addresses. If

your organization must receive mail from senders who use IPv6 addresses, create a dedicated Receive connector that restricts the remote IP addresses to the specific IPv6 addresses that those senders use. If you configure a Receive connector to accept anonymous connections from unknown IPv6 addresses, the amount of spam that enters your organization is likely to increase."

The expectation that adoption of the new protocol will exacerbate spam problems sets a standard of best practices, which reduces incentives to fully implement IPv6, even when it is bundled, paid for, and ready to deploy at a click of a mouse. The recommendation that that IPv6 channels should be restricted is not specific to this particular product, and is indeed common. For instance, according to the database maintained by the United States Computer Emergency Readiness Team (US-CERT), IPv6 specific security and spam related vulnerabilities have been detected in many other networking applications. [25] The data also reveal that most specific issues have been patched in a timely manner by vendors, but that whitelisting only IPv6 implementation policies remain a realistic approach to mitigating perceived risks of spam. The unavailability of IPv6 Domain Name System Real-time Black Lists (DNSRBLs) represents a common reason behind such advice.

Unlike end users that may not have the technical capacity to chose the protocol in use, managers of downstream networks are concerned with security issues. While the adoption of OSs that support the standard by default can create demand necessary to invest in IPv6 infrastructure, a heavy reliance on whitelisting reduces the probability of mass diffusion. Unless vendors of this class of products bundle antispam mechanisms that reduce the need for this type of implementation policy, the configuration of downstream servers can break the link between demand by end users and supply of IPv6 infrastructure by upstream ASs.

(Semi) Autonomous Systems: End users and downstream providers have some capacity to influence decisions about the implementation of the new protocol. Both classes of participants typically purchase bundled operating systems and applications. Hence, they do not actively choose the Internet Protocol in use, but have some degree of freedom in making decisions shaping IPv6 adoption. In contrast, upstream entities that connect enterprises, public institutions, and sometimes end users, have strong economic interests in the IP choice. In the context of the diffusion problem detailed in the first section, upstream providers are also likely to have a larger number of linkages to others (larger *d*'s). Consequently, decisions by this class of operators are likely to be of particular relevance in the diffusion of IPv6.

In the early days of the spam problem, this class of operators generally did not interfere with the flow of messages, and left the issue to downstream administrators and end users. With the growth of spam volumes since the early 2000s, network providers have increasingly had to start filtering incoming flows based on DNSRBLs and systems that try to assess the reputation of senders. Although not very accurate in detecting spam (50-60%) relative to content scanning and classification, with the growth of spam this approach to spam filtering has become an integral part of the messaging infrastructure of service providers. [26] Centralized reputation systems also allow administrators to discard some portion of the noise at the edge of their network, reducing its costs on operators in terms of hardware, bandwidth, and storage. [27]

Given these advantages, network operators have invested heavily in reputation based filtering technologies in the past few years. For a number of reasons detailed below, IPv6 reduces the expected payoffs from the investments in reputation and authentication based mechanisms to mitigate the costs of spam. In the presence of sunk costs, decisions about IPv6 infrastructure investments across ASs are likely to be lumpy. Replacing these systems with spam classification mechanisms that do not rely on the identity of senders will be required to deal with this microeconomic problem.

Spammers: Some observers have argued that diffusion of IPv6 may enhance the capacity of

reputation systems under certain conditions. Davis et al. (2006) point out that if it was possible to create ownership in IPv6 address space, the resulting system would allow for the assignment of permanent addresses to all possible machines for the duration of their life. [28] The development of such a property system would make reputation more credible than when spammers can easily forge sender addresses and cloak their identities. They further argue that IPv6 adoption will limit the well known false positive problem associated with the use of reputation and blacklists as an instrument of spam control. Even if it was possible to assign static IP addresses and create a property regime, this view appears spurious for a number of reasons.

Increased use of blacklists and reputation systems since the mid-2000s has stimulated spam system developers to deploy a wide array of BGP spectrum agility techniques and fast-flux networks of spam robots. [29, 30] With these innovations, spammers can send just one message from millions of different machines, hence reducing the informational value of sender reputation to the spam control process. The relatively low accuracy of reputation systems highlights difficulties created by these techniques under IPv4. The large address space under IPv6 and the assignment of IPv6 addresses to various kinds of machines increases the range of opportunities for engaging in one shot agility techniques with the use of large distributed robots currently available to spammers.

In addition, the widespread use of permanent addresses can illicit a second undesirable response from spammers. If such a property regime is imposed, for instance through a global mandate, providers of spam systems will have incentives to invest in techniques aimed at hijacking the identity of senders with good or neutral reputations. Implicit coordination by large network operators in the form of reputation feedback loops has already motivated the deployment of some reputation hijacking techniques under IPv4. In fact, these techniques partly explain the false positive problem with DNSRBLs and reputation based filtering that Davis et al. (2006) hope to alleviate with a property regime. The hypothesis that IPv6 will limit the spam problem and enhance the effectiveness of reputation systems is misguided because it does not consider the current technological frontier and the innovative nature of the market for spam production software.

Davis et al. (2006) further contend that the availability of a near infinite address space will increase the costs of spamming because it will require greater resources to ping and probe different systems. This conclusion similarly does not account for the well-documented capacity of spammers to innovate in response to emerging challenges. There are really no significant technical barriers for spammers to implement more refined methods to search for vulnerable ports, collect information about potential targets, and organize their campaigns. The difficulties posed by the large IPv6 address to spammers consequently appear trivial.

As far as the authors are aware at the time of writing this paper, no IPv6 specific worm or spam botnet has been deployed yet. This observation likely reflects the current low penetration rates of the protocol, hence the absence of a viable market for which to develop malware. However, attackers can solve the so called missing market problem much more easily than network operators by employing a wide range of available techniques for active scanning of computers that are easy to contaminate. Although these techniques are relatively inefficient, it is important to assume that they will be adapted to an IPv6 environment if the new protocol offers a viable channel for increasing the spammers' conversion rate. Bellovin et al. (2006) for instance describe how a two stage search process (wide/local area) that takes advantage of the neighbor discovery logs on an IPv6 host can be used to collect valuable information about machines on a LAN. [31]

In addition to their ability to use active and smart scanning, computers in an IPv6 address space are likely to live near each other, much like people agglomerate in cities. This means they will be provided with addresses that are adjacent to each other, since this is the easiest way to assign the next free address by the

service provider. This structural feature of the new protocol enhances the capacity of spammers to deploy worms that collect information required to build mailing lists. The same worms can be employed to distribute armies of robots that are necessary to bypass reputation systems with one shot games.

Most computers in the world are already protected from random scanning by NATs. As a result, spammers have already shifted away from active scanning. In fact, they are increasingly deploying techniques that let their pray come to them instead. It is sufficient for instance to place some malicious code on servers that are commonly visited, and collect relevant information about the location of entities using the server. WWW servers also can be used to similar ends. Since WWW and mail servers provide useful services, their operators tend to advertise their location, which saves the spammers the trouble of active scanning. Starting from a few specific servers, attackers can identify the address of the next linked group of servers and so on. Passive scanning can also be implemented by listening and collecting the necessary information about targets infiltrating routers of big service providers or by using BGP to redirect part of the traffic to the spammer.

IV. Summary and implications

The costs of spam in terms of scarce resources of end users, as well as the infrastructure of operators are well known. This paper hypothesizes that the costs of spam for the growth of the Internet are likely to be more pernicious than previously believed. This is because the prospects of increased spam can reduce incentives for the diffusion of efficiency enhancing technologies. We documented this hypothesis in the case of IPv6, a bundle of standards that appear necessary for increasing access and reducing the fragmentation of the system through NATs.

The first part of the paper provided an overview of the motivation for IPv6 adoption, as well as the state of supply and demand for the standard. Economic theory suggested that decisions by large network operators, with the largest number of linkages, are likely to be central in the diffusion of the standard. It also helped explain why bootstrapping efforts and partial mandates have failed to increase penetration rates, despite widespread acceptance of the costs of IPv4 address exhaustion. The second part of the paper detailed the likely impact of technical features of IPv6 on decisions by end users, downstream and upstream network providers, and of course, spammers.

The analysis suggests that IPv6 implementation will erode the capacity for spam mitigation based on DNSRBLs and mechanisms that aim to assess the reputation of senders. While this insight may appear trivial to some observers, establishing precisely why it is the case is important for proponents of IPv6 and designers of antispam systems.

Given the interdependence of decisions by Autonomous Systems, the possibility of a multiple equilibrium problem in the diffusion of IPv6 is realistic. Failures of regulatory mandates and bootstrapping efforts around the world to increase penetration rates in the past few years support this hypothesis. Overall, IPv6 appears to alter the relative capacity of attackers and defenders in a manner that increases the likelihood of a stable, but inefficiently low equilibrium penetration rates. Facing sophisticated spammers and reliant on reputation systems, network operators are reluctant to invest in IPv6 infrastructure.

The analysis also described another specific channel through which risk aversion to spam is likely to influence the IPv6 diffusion process. The transition from Windows XP to Vista, where both standards are functional by default, might generate sufficient end user demand to stimulate investment in IPv6 specific infrastructure. However, the perception that full IPv6 functionality results in an erosion of antispam capabilities forces administrators to restrict access and apply whitelisting only policies. At the limit, the combination of address exhaustion and this policy will result in the creation of networks that deploy IPv6 internally, but continue to employ IPv4 with the outside. For instance, an operator that needs IP addresses to control a large number

of military vehicles or cable modems can use this strategy to deploy IPv6 without exposing her network to increased spam.

For organizations that place a high priority on security, such as a military or a bank, protecting internal IPv6 networks through Network Address Translation and masquerading techniques may be justified. A NAT would limit the capacity of infiltrators to actively scan for port and information about their potential targets. Replication of this strategy to the standard may not be efficient for network providers where the level and quality of end user access is the priority. IPv6 to IPv4 translation may deal with the address exhaustion problem within some operators, but will not improve the level of IPv6 connectivity across ASs. Since decisions by network operators are likely to reinforce each other, this trend should be a particular worry for proponents of the new protocol.

A number of other possible solutions to the problem detailed here have been proposed. Otis (2008) proposes the adopting stronger authentication protocols and reputation feedback loops between large operators. Given the expected difficulties with DNSRBLs and reputation based mechanisms, Otis (2008) suggests the complementary use of sender authentication protocols such as DKIM with IPv6. [32] However, the use of DKIM and SPF creates new challenges for network operators and angles of attack for spammers, which explains why they have not received widespread support. [33] Andersen et al. (2008) describe an Accountable Internet Protocol (AIP) which in theory could address DNS spoofing and route hijacking that reduce the effectiveness of protocols such as DKIM or SPF. [34] While intriguing, their proposal is not likely to solve the basic problems associated with classifying messages based on information about purported senders that reduces incentives to invest in IPv6.

Consequently, diffusion of IPv6 appears to require a shift to spam filters that focus more on the content of messages rather than the reputation of senders.

The capacity of spammers to engage in one shot spectrum agility techniques and hijack good/neutral reputations means that decisions by this class of filters are likely to become increasingly unreliable. Expectations of lower accuracy and increasing false positives can stimulate excessive levels of whitelisting only policies, or IPv6 “lite” implementations. This will not increase IPv6 connectivity across ASs or stimulate incentives for its diffusion.

References

[1] Metrics reports by Messaging Anti-Abuse Working Group (MAAWG) provide some useful statistics. http://www.maawg.org/

[2] Cormack and Lynam (2007) On-line Supervised Spam Filter Evaluation, ACM Transactions on Information Systems 25, 3.

[3] ITU Study on the Financial Aspects of Network Security: Malware and Spam. Final Report, 2008.

[4] Alperovitch and Krasser (2007) A Taxonomy of Email Reputation Systems, ICDCS Workshops.

[5] Kimakova and Rajabiun (2008) Multilayer Filtering: The Dangerous Economics of Spam Control. The Proceedings of the MIT Spam Conference.

[6] Potts (2007) Worldwide IPv6 Initiatives, Interworking, Santiago, Chile.

[7] Kearns (2005) Economics, Computer Science, and Policy. Issues in Science and Technology.

[8] http://tools.ietf.org/html/rfc2460

[9] http://www.caida.org/research/policy/

[10] IPv6 @ Comcast. http://www.ripe.net/ripe/meetings/ripe-54/presentations/

[11] Report to congressional requests on IPv6 by United States Government Accountability Office (GAO) on IPv6, May 2005.

[12] http://tools.ietf.org/html/draft-durand-dual-stack-lite-00

[13] http://ipv6.nlsde.buaa.edu.cn/

[14] Gunderson (2008) Google global IPv6 statistics: Measuring the current state of IPv6 for ordinary users. RIPE 57, Dubai.

[15] http://marketshare.hitslink.com/report.aspx?qprid=11&qpdt=1&qpct=4&qptimeframe=M&qpsp=97&qpnp=25h

[16] Elmore, Stephens, and Camp (2008) Diffusion and Adoption of IPv6 in the ARIN Region. *Workshop on the Economics of Information Security* (WEIS).

[17] Jackson and Yariv (2007) Diffusion Behavior and Equilibrium Properties in Network Games. American Economic Review, 97, 2.

[18] Ozment. and Schechter (2006) Bootstrapping the Adoption of Internet Security Protocol.. *Workshop on the Economics of Information Security* (WEIS).

[19] Loder, Van Alstyne, and Wash (2004) Information Asymmetry and Thwarting Spam. http://web.mit.edu/marshall/www/home.html

[20] Xiao, Liu, Guo, and Xu. (2009) Modeling the IPv6 Internet AS-level Topology. Physica A 388:529-540

[21] Kanich, Kreibich, Levchenko, Enright, Savage, Voelker, and Paxson (2008) Spamalytics: An Empirical Analysis of Spam Marketing Conversion. ACM Conference on Computer and Communications Security.

[22] Androutsopoulos, Magirou and Vassilakis (2005) A Game Theoretic Model of Spam E-Mailing. CEAS 2005.

[23] Reshef and Solan (2006) The Effects of Anti-spam Methods on Spam Mail. CEAS 2006.

[24] O'Donnell (2008). When Malware Attacks (anything but Windows). The Proceedings of the MIT Spam Conference.

[25] http://www.us-cert.gov/

[26] Results of Anti-spam Solution Testing. Opus One, 2007.

[27] Alperovitch, D.; Judge, P.; Krasser, S. A taxonomy of email reputation systems. ICDCS Workshops, 2007.

[28] Davis, Spong, and Bray (2008) Keep Your Spam Off Me! http://www.ent.ucf.edu/

[29] Ramachandran and Feamster. (2006) Understanding the Network-Level Behavior of Spammers, SIGCOMM 06, Pisa, Italy.

[30] Nazario and Holtz (2008) As the Net Churns: Fast-Flux Botnet Observations, Malicious and Unwanted Software. MALWARE 2008.

[31] Bellovin, Cheswick and Keromytis (2006) Worm Propagation Strategies in and IPv6 Internet. LOGIN, 31, 1.

[32] Otis (2008) IETF, DKIM Working Group, Draft 3: DKIM Author Domain Signing Practices (ADSP) Security Issues.

[33] Ostrihon and Rajabiun (2008) The Robustness of New Email Identification Standards. The Proceedings of the Virus Bulletin Conference, Ottawa.

[34] Andersen1, Balakrishnan, Feamster , Koponen , Moon , and Shenker (2008). Accountable Internet Protocol (AIP). SIGCOMM 08.