MANAGING REPUTATION IN AN IPv6 WORLD

AUTHOR:
Alec Peterson, VP of Technical Services, Message Systems

CO-AUTHOR:
Robert Marchi, Senior Solutions Architect, eleven

PREFACE:
Alec Peterson, VP of Technical Services, Message Systems
Table of Contents

2 PREFACE

3 EXECUTIVE SUMMARY

3 WHAT’S DRIVING IPv6

4 THE IPv6 REPUTATION CHALLENGE

4 PROPOSED IPv6 REPUTATION APPROACH
   4 Overview
   5 IPv6 Policy Announcements
   6 Policy Query Technology

7 CONSUMER APPLICATIONS OF IPv6 ASSIGNMENT POLICY

8 IPv6 ASSIGNMENT POLICY ANNOUNCEMENT RISKS

8 CONTENT APPLICATIONS ISSUES

9 CONCLUSION
Preface

As Internet Service Providers (ISP) prepare to deploy IPv6, many have expressed concern about the potential impact on how abuse is perpetrated and defended against. In the realm of electronic messaging security, IP reputation is a critical element in the defense against abuse. Various providers track the reputation of individual IP addresses and make this information publicly available. Many email receivers opt to limit or refuse traffic based on these lists.

The stakes are high since abuse represents approximately 90% of all email traffic. Today, ISPs use IP reputation to stop between 60–90% of all abusive inbound traffic at the gateway. Some content scanning applications also factor IP reputation into the spam score they assign to messages. In short, IP reputation is an integral part of any comprehensive messaging security deployment.

IPv6 expands the Internet address space from 4 billion addresses to an incomprehensible $3.4 \times 10^{38}$ addresses. Such exponential growth in IP addresses makes it infeasible to track reputation using the conventional methods employed with IPv4. Furthermore, even were it possible to track that many different addresses, such a database would be essentially useless because an individual subscriber has access to such a vast amount of address space. This then begs the question:

How to track IPv6 reputation in a meaningful way?

This paper proposes an answer to that question — an approach for tracking IP reputation in the expanded address space of IPv6. We believe that it’s doable, despite what others have said, and see this paper as the starting point for discussions within the ISP community on how it might be achieved. We invite others to join in this discussion.

— Alec Peterson, VP of Technical Services, Message Systems
Executive Summary

When ISPs begin deploying IPv6, Regional Internet Registries (RIRs) will allocate them an address space that is orders of magnitude larger than the entirety of IPv4 address space. In turn, ISPs will assign a very large block of address space to each of their customers. Individual subscribers will have quintillions of IP addresses available to them (that's thousands of quadrillions, or millions of trillions). This means the Internet address space (possible Internet addresses) will suddenly grow at exponentially higher rates than those of today. We submit that the growth of the number of entities using the Internet will remain essentially the same. Thus, the industry should shift to tracking the reputation of these entities rather than individual IP addresses. This paper details the perceived and likely impact of IPv6 on techniques for combating messaging abuse, including:

- **Challenges created by IPv6.** The new address space creates the challenge of tracking an exceedingly high number of IP addresses.

- **A proposed method for tracking IP reputation.** By tracking blocks of Internet addresses assigned to subscribers rather than individual IP addresses, reputation providers will be able to track the number of entities comparable to those tracked today.

- **Proposed use and implementation of IPv6 Assignment Policy Announcements.** Publication of ISP IP assignment policies will allow reputation providers to know how to identify entities on an ISP's network.

What’s Driving IPv6

The Internet has grown beyond what anybody involved in its early development could have imagined. Who would have thought that a network developed in the 1970s for government and education entities to communicate would have become an indispensible utility touching every aspect of our lives?

One of the things that the original architects of the Internet didn’t anticipate was that Internet Protocol (IP) addresses — the essential way that electronic devices communicate with each other — would become exhausted. Back in the 1970s, the 4 billion addresses available under the IP address assignment scheme known as Internet Protocol version 4 (IPv4 for short) seemed more than ample. But just months ago (early 2011), the last blocks of IPv4 address space were released from the Number Resource Organization to the Regional Internet Registrars.

Fortunately, the groups responsible for defining standards for the Internet saw this day coming long ago, and began developing a new version of the Internet Protocol, called Internet Protocol version 6 (IPv6). They did not pursue version 5 since it was experimental and never saw wide deployment. As opposed to IPv4 which has roughly $4.2 \times 10^9$ addresses, IPv6 has roughly $3.4 \times 10^{38}$ addresses — larger than the number of particles believed to exist in the universe. Such a vast address space means that once the Internet and devices support IPv6, there will be no issue with the number of available addresses for a very, very long time.

Unfortunately, the sheer volume of the IPv6 address space introduces challenges that do not exist in an IPv4 network. For example, many techniques used to track the reputation of an IP address do so based on the entire IPv4 address. Since there are ‘only’ 4 billion possible IPv4 addresses, this is not an unreasonably large number of addresses to account for. Taking into account the large amount of available IPv6 address space, this approach is not feasible with IPv6 addresses.
The IPv6 Reputation Challenge

Spam makes up about 90% of all email traffic on the Internet. In order to protect network resources, large volume receivers such as ISPs use IP address reputation (IP reputation) lists to drop connections from malicious sources prior to receiving message content. IP reputation and other techniques allow receivers to eliminate 60–90% of abusive inbound traffic. Once the message data is received, content applications are used to evaluate the message content. Many content applications use message header and IP address information to rate a message. Thus, IP reputation is a critical component of any effective anti-abuse implementation.

Many experts have asserted that for inter-server traffic, IPv6-enabled servers should move to a whitelist model, and away from the current model of allowing essentially any terminal on the Internet to connect to what amount to infrastructure servers. This is an interesting approach, but it does not at all diminish the need for a multi-faceted approach to handling IPv6 traffic. Fundamentally, services in an IPv6 environment need to be divided into infrastructure services on the one hand, and customer premise equipment-facing services on the other. Any service that is intended purely for inter-carrier communication (such as the ‘inbound’ SMTP servers for an ISP’s email messaging platform) realistically only need to accept connections from ‘known’ clients. However, many servers must provide access to essentially any IPv6 address (such as HTTP/HTTPS servers and ‘outbound’ SMTP servers just to name a few). It is these servers that must be able to quickly and cost-effectively track reputation of any IPv6 address, as they do today with IPv4 addresses.

Proposed IPv6 Reputation Approach

Overview

While IPv6 changes many aspects of the infrastructure of the Internet, it does not change the fundamental nature of the business relationship between ISPs and their subscribers. Said another way, today’s subscriber who gets a block of IPv4 /32 (1) or /28 (16) addresses will be the same subscriber who tomorrow gets an allocation of /64, (1.8x10^19) or /56 (4.7 x 10^21). Even though a subscriber may have a very large number of available addresses, the adoption of IPv6 will not result in a massive increase in the number of actual subscribers or even the number of IP addresses in use on the network. In other words, IPv4 reputation has often been a way of tracking subscriber reputation, where one subscriber equals (roughly) one IP address. IPv6 obviously breaks that model, but ultimately there still will be one IPv6 prefix assigned to a single subscriber.

Thus, the challenge then becomes how to move from a model of mapping a single IP address to a subscriber to a model of mapping an IP prefix to a subscriber. This means considering all IP addresses within a given range of addresses together as one, as opposed to the current simplistic mechanism of considering each IP address individually with no aggregation of results. Implementation of such an approach requires an analysis of what data is immediately available today as a priori knowledge, versus what additional data is needed.

The only thing the receiver of an SMTP connection (an SMTP server) knows about the client initiating a connection at the time of connection creation is the source IP address, and source TCP port number. The source TCP port number is random, and for the purposes of this discussion generally meaningless. Based on the source IP address other information may be learned today; specifically a reverse DNS query may be performed to attempt to extrapolate a hostname associated with the IP address.
In an IPv6 world this alone introduces some new complexities; as with the extremely large amount of host-address space available to subscribers, an intelligent method of returning the same reverse DNS value for all hosts within a /64 (for example) will need to be developed. Such a method would also be helpful for other aspects of this suggested approach, and will be discussed later.

Given a source IPv6 address (and potentially a hostname), one cannot yet determine how to aggregate addresses for a specific client. This is because there are a variety of competing proposals regarding the size block that will be allocated to a subscriber. Without going into details on the pros and cons of the competing proposals, it is generally feasible that a service provider may perform allocations to its clients with a prefix length anywhere from /48 to /64. Prefix length will likely be on a nibble boundary (/64, /60, /56, /52, /48) but nothing will prevent a service provider from arbitrarily assigning a /61, for example. Assuming a receiver knows what prefix length a service provider used for a given client, this now means IPv6 reputation is fairly trivial to associate a source IPv6 address with a client, which completes the loop that IPv4 reputation completes.

From a policy standpoint, there is still a remaining gap in this proposal. What incentive does a service provider have to announce its assignment policy at all? If this is the case, this proposal suggests assigning a default aggregation policy of /48 to all IPv6 allocations which do not announce an assignment policy. As /48 is the shortest direct assignment size allowed, such a default policy is both reasonable and prudent. However, this creates some risk for the service provider. If the service provider assigns smaller blocks to their subscribers, the actions of one IPv6 subscriber could negatively impact the reputation of another. This in turn has the potential to create a customer service issue for the provider. Subscribers using an address within a blacklisted prefix would be unable to send messages and customer service calls would ensue. ISPs are therefore incentivized to announce their ‘real’ IPv6 assignment policy to ensure the actions of one IPv6 subscriber do not negatively reflect upon the reputation of another.

**IPv6 Policy Announcements**

The variable nature of possible prefix assignment lengths necessitates a mechanism for service providers to distribute their assignment policy for a given block of address space. DNS has historically been used for distribution of such policy information (note that SPF and SenderID are essentially communicating policy regarding desired email reception policy). Therefore, applying the DNS infrastructure to this problem leverages a pre-existing and well-proven technique.

A closer examination of the SPF and SenderID use cases mentioned above reveals that they are based on ‘forward DNS’ infrastructure, in other words DNS based on a domain name (such as messagesystems.com). Distribution of IP-based policy will not work in such an environment, because at connect time a server only knows the IP address, and has no information about what domain to query for IP assignment policy. Fortunately, infrastructure already exists that maintains a necessary IP-based hierarchy, in the form of the IPv6 reverse DNS infrastructure (the ip6.arpa domain, the IPv6 version of in-addr.arpa used for IPv4).

Reverse DNS lookups themselves use a special DNS record type called a ‘PTR’ record. Therefore, other record types sharing the same DNS name space will not conflict. Ultimately a technique such as this should have its own record type, but for the sake of a streamlined process to deployment the TXT record type is an appropriate one to use. TXT records are already used in ‘forward DNS’ infrastructure for DKIM, SenderID and SPF data. Each of those record formats is marked with a version tag to avoid confusion with other uses of the TXT record. Using a similar technique for IPv6 Assignment Policy
Announcements will ensure this use is consistent with other uses. For example, a service provider who has been allocated 2001::1200/40 and has a /56 assignment policy within this block would place the following record in its reverse DNS zone:

`*.2.1.0.0.0.0.0.1.0.0.2.ip6.arpa. IN TXT "v=IPV6POL1 a=56 m=40"`

The tags defined therein are as follows:

- **v**
  - version string (text, REQUIRED). String comparison (so IPV6POL1 is different from IPV6POL1.0). Current value MUST be ‘IPV6POL1’.

- **a**
  - assignment mask length (integer, 48–64 inclusive, OPTIONAL, default 48). For IPv6 addresses which caused this record to be returned, consumers of these data may reasonably assume all traffic received from addresses encompassed by an aggregate based on this mask are from the same subscriber.

- **m**
  - policy aggregate mask length (integer 0–64 inclusive, OPTIONAL, ignored if < value of the ‘a’ tag). Describes the size of the aggregate to which the assignment mask length applies. Consumers of these data may cache query responses and apply the same assignment mask length policy to all IPv6 addresses within this aggregate.

SECURITY NOTE: This could be used to announce a policy for more address space than a service provider controls (by specifying an m= value which is shorter than the actual size of the allocation that the service provider has received). Implementers MUST perform a query against the full aggregate alone to verify the data is the same before caching the response.

Policy Query Technology

Queries done based on IPv4 addresses (be they reverse DNS, RBL lookups, or any other use) were generally done in a very simplistic manner. Specifically, for a given IPv4 address (say 1.2.3.4) a query against an RBL would be done against 4.3.2.1.myrbl.com for the appropriate record type. The relatively small scope of the address space combined with the fact that a terminal could only use one (or a relative handful) of addresses significantly limited both the number of distinct queries that a single terminal could force a server to make, as well as the amount of data it would need to cache.

IPv6 introduces some significant challenges to perpetuating that model. Even if one were to ignore the challenges of reputation relating to IPv6, there are significant issues facing those who just perform simple reverse DNS queries. Consider an abusive sender attached to a subscriber connection which has been allocated a /64 of IPv6 address space. Such an abusive sender has $1.8 \times 10^{19}$ available addresses and therefore could change his IPv6 address several times per second while making connections to various servers. If such servers were performing reverse DNS validation of each connection, the result would be a significant increase in not only the amount of DNS data the servers being connected to must track and cache, but also the amount of traffic that the DNS servers hosted by the subscriber’s service provider must process. Such load is significantly above what is possible to generate in an IPv4 world simply because of the smaller set of distinct reverse DNS queries that can be generated by a single terminal. Caching techniques cause repeated connections from the same IPv4 (or IPv6) address to be suppressed in the event a locally cached answer is located.

---

1 This technique uses a DNS wildcard, which some operators may prefer not to use. Tools such as `rbldnsd` can also implement behavior such as this, but currently `rbldnsd` does not support IPv6. `rbldnsd` should be updated to support IPv6 in the near future, in the mean-time the wildcard approach works for our purposes.
The problem of abusing the reverse DNS infrastructure in an IPv6 world is beyond the scope of this document, but the issue is illustrated here to highlight a fundamental shift in DNS query methodology that is required as part of the migration to IPv6. Fortunately, a solution to this abuse concern is not difficult to envision. Platforms implementing this technique must be able to intelligently cache and aggregate results that are repeatedly returned. For instance, take the example described above where 2001::1200/40 uses a prefix assignment length of /56. A client connects from a source address of 2001::1234:5678:0123:4567:89ab:cdef, which results in the following DNS policy query:

f.e.d.c.b.a.9.8.7.6.5.4.3.2.1.0.8.7.6.5.4.3.2.1.0.0.0.0.1.0.0.2.ip6.arpa TXT

As defined above, the wildcard DNS entry will return the TXT record that was defined, specifically:

“v=IPV6POL1 a=56 m=40”

When a client connects, the only thing a server knows is the source IP address. Thanks to the above DNS query, we now also know:

1. The client connecting was actually assigned a /56 prefix, specifically 2001::1234:5600/56, which means that all connections coming from that prefix can be assumed to be from the same subscriber.

2. The assignment policy applies to a block of length /40 (as per the ‘m’ field).
   That means we may avoid performing any further DNS queries for any IPv6 addresses we receive connections from that are within the 2001::1200/40 range of address space.

Platforms implementing this technique should use the information gleaned above to cache IPv6 assignment data appropriately. Failing to perform such caching will create significant denial of service vectors, both against the platform itself as well as against third party reverse DNS servers. Furthermore, caching for negative DNS queries is just as important (in other words, DNS queries that result in no IPv6 Assignment Policy record being returned). As the default assignment policy has already been defined as being 48 bits long, negative caching based on that value is both correct and prudent.

**Consumer Applications of IPv6 Assignment Policy**

Both commercial entities and non-profit groups maintain global reputation lists today based on raw data collected from service providers across the Internet. These lists are generally kept as data elements in log files detailing valid and abusive traffic, and keyed by source IP address. This collection of data would be essentially the same for an IPv6 reputation list, except the source IP addresses would be IPv6 instead of IPv4.

Techniques will need to evolve when such data is aggregated into both an internal database for analysis, as well as the summarized reputation list that is provided back to ISPs and other users of the reputation list. Today most global reputation providers evaluate individual IPv4 addresses autonomously, relying on common patterns among adjacent IPv4 addresses to create sensible aggregates. The sheer size of available addresses in an IPv6 environment will make such post-facto aggregation extremely difficult to achieve. Utilizing the query techniques described earlier in this document, a global reputation provider may learn the a priori aggregate size for a given address, and subsequently continue to provide effective reputation for IPv6 addresses.
Consumers of global reputation lists will also need to exercise care when querying the list. If DNS is involved in any stage of the query process, the caching risks described above for policy queries also apply to queries against the reputation lists themselves. Global reputation providers may enable a similar aggregation scheme as enabled by the ‘m’ tag in the IPv6 Assignment Policy record when returning the TXT record for a lookup query. While global reputation providers are responsible for how their lists are structured and queried, one way or another, this challenge will need to be addressed (as will how to address negative hits).

Real-time local reputation is tracked in a similar manner to global reputation. Raw local events are collected, generally based on a single IPv4 address, stored in an aggregate form, and queried at a later time to quickly determine the reputation of a connection peer. In an IPv6 environment, source IP addresses must first be rolled up into aggregates described by assignment data learned from DNS queries explained above. Furthermore, subsequent reputation queries against said aggregated data must be performed such that the aggregated data is properly evaluated.

**IPv6 Assignment Policy Announcement Risks**

Risks related to this proposal revolve around ISPs announcing a misleading assignment policy to attempt to ‘game’ receiver reputation filters by claiming to assign a longer prefix (for example, /64) than is really assigned to a given client (for example, /48). In this scenario, a subscriber with a /48 would actually have 65,535 times as much room to manipulate a reputation filter for a service provider. Such is the risk either for the direct recipient of address space from a registrar (otherwise known as a ‘registrant’) or potentially a service provider further downstream of the registrar similarly providing such false information. This risk is not significantly different than the risk today of a malicious subscriber being assigned multiple distinct IPv4 addresses. The same challenge of figuring out how to map the addresses together is similar to a malicious IPv6 subscriber announcing an incorrect policy. Therefore, the risk does not significantly change the model from how it exists today.

**Content Applications Issues**

As malicious senders seek to avoid both protocol-level and content-level (data phase) filtering applications, they have continually evolved to introduce more randomization in the content. This randomization has increased the challenge of analyzing the content. To combat this, some applications now opt to analyze portions of the message header and envelope information. As IPv6 is introduced, this will impact the analysis of the header information. The challenges in tracking the IP reputation discussed above may also apply to content applications.

If the application ascribes a reputation to IP addresses, the application provider must deal with the new address space as well. Content provided to the application will contain source IP and purported domain information, but this will not include information the assignment policy looked up during the connection phase of the SMTP transaction. Typically content applications off load the analysis work to a central data center so that data can be compared with traffic from all their instances. Depending on their use of IP information, they may also require similar tracking of the senders’ allocation policy in order to properly manage the grouping of IP entities. Receivers deploying content applications should closely review their vendor’s implementation to gain an understanding of the impact of IPv6 on their application.
Conclusion
The sunset of IPv4 and subsequent deployment of IPv6 introduces a wide variety of challenges, of which reputation tracking is one of many. As a commonly held public resource, the Internet will continue to be a target-rich environment for bad actors, and we can expect spam and other kinds of abuse to pose problems for ISPs (and everyone who uses the Internet) well into the future. That is not to say that the transition to IPv6 necessarily presents a golden opportunity for perpetrators of abuse. With planning and foresight, these challenges can be addressed.

A tremendous amount of attention is currently being paid to how to upgrade the infrastructure of various networks to support IPv6, which is clearly a dependency upon which everything else related to IPv6 relies. Nonetheless, if sufficient attention is not paid to issues such as IPv6 reputation tracking, we will find ourselves in a very difficult position of having to hastily solve a very complex problem with a sub-optimal technique. Looking ahead to the challenges we are likely to face will only improve the chances of a smooth and efficient migration to IPv6.

WHAT MESSAGE SYSTEMS OFFERS
Message Systems Momentum is the industry’s only true message management platform. Serving as a controlling gateway, intelligence router and advanced MTA, Momentum is designed as the central communications hub for connecting business processes, data sources, applications and systems. Momentum’s open, scalable framework allows ISPs, carriers and enterprises to easily handle carrier-grade message loads, and integrate email and mobile messaging into their business processes in innovative new ways. Momentum has powerful capabilities for controlling message flow and processing multi-channel communications based on custom business rules. Momentum also offers full IPv6 support, providing the technology foundation for carrying out a successful reputation-tracking strategy in the new IPv6 environment. For more information, go to messagesystems.com or call +1-410-872-4910 (international) or 877-887-3031 (toll free USA).

WHAT ELEVEN OFFERS
eleven’s eXpurge content filtering application employs sophisticated fingerprinting analysis algorithms to make definitive classifications of email content. By tracking the distribution of content across over 47,000 domains around the world, eXpurge rapidly detects new threats in real time. eXpurge utilizes the source IP address to analyze how content is replicated across the Internet. However, rather than ascribing a reputation to IP addresses, eXpurge associates the content distribution with source distribution. This approach is agnostic to the IP address space being analyzed. Thus, eXpurge is ‘IPv6 ready’ today. For more information, visit www.eleven.de.
ABOUT MESSAGE SYSTEMS
The industry leader in messaging technology, Message Systems offers a family of software solutions and services that addresses the digital communications needs of today’s most innovative companies. Telecommunications carriers, ISPs, marketing services providers, cloud computing firms and social media companies rely on Message Systems software to power the message-based communications driving their critical business initiatives. Message Systems solutions get billions of unique messages to the right place at the right time every day through the full range of channels: email, SMS text, MMS messaging and more. Founded in 1997, the company is headquartered in Columbia, Maryland.

ABOUT ELEVEN
eleven is a leading email security provider based in Germany. Its eXpurgate technology, which is unique worldwide, offers a spam filter and email categorization service that protects the user reliably from spam and phishing, detects potentially dangerous email and can distinguish between individual messages and any kind of mass email. eXpurgate also offers numerous virus protection options and a powerful email firewall. Over 45,000 companies of all sizes use eXpurgate to check and categorize more than a billion email messages every day. Customers include Internet service providers and telecommunication carriers such as T-Online, O2, Vodafone and freenet as well as many well-known companies and public institutions, including Air Berlin, the Federal Association of German Banks, DATEV, the Free University of Berlin, Landesbank Berlin, Mazda, RTL, ThyssenKrupp and Tobit Software AG.