Information-Security

IPv6-Reconnaissance

Internet-weite Analyse von ICMPv6

Diplomarbeit

zur Erlangung des akademischen Grades

Diplom-Ingenieur/in

eingereicht von

Florian Holzbauer, BSc

1810619809

im Rahmen des

Studienganges Information-Security an der Fachhochschule St. Pölten

Betreuung

Betreuer/in: Dipl.-Ing. Peter Kieseberg

Mitwirkung: Dr. Johanna Ullrich

St. Pölten, 22. Juni 2020 _____ _____

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

*

Fachhochschule St. Pölten GmbH, Matthias Corvinus-Straße 15, A-3100 St. Pölten,T: +43 (2742) 313 228, F: +43 (2742) 313 228-339, E:office@fhstp.ac.at, I:www.fhstp.ac.at

Ehrenwörtliche Erklärung

Ich versichere, dass

- ich diese Diplomarbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich sonst keiner unerlaubten Hilfe bedient habe.
- ich dieses Diplomarbeitsthema bisher weder im Inland noch im Ausland einem Begutachter/einer Begutachterin zur Beurteilung oder in irgendeiner Form als Prüfungsarbeit vorgelegt habe.
- diese Arbeit mit der vom Begutachter/von der Begutachterin beurteilten Arbeit übereinstimmt.

Der Studierende/Absolvent räumt der FH St. Pölten das Recht ein, die Diplomarbeit für Lehre- und Forschungstätigkeiten zu verwenden und damit zu werben (z.B. bei der Projektevernissage, in Publikationen, auf der Homepage), wobei der Absolvent als Urheber zu nennen ist. Jegliche kommerzielle Verwertung/Nutzung bedarf einer weiteren Vereinbarung zwischen dem Studierenden/Absolventen und der FH St. Pölten.

Ort, Datum *Unterschrift*

Kurzfassung

IPv6 ist ein fundamentaler Baustein des Internets, das seinen Vorgänger IPv4 auf Netzwerkebene ersetzen soll, um das Problem der Adressknappheit ein für alle Mal zu lösen. Obwohl das neue Adressformat bereits 1998 in RFC2460 spezifiziert wurde, erfolgt die Umstellung nur in langsamen Schritten. Im November 2019 allokierte die RIPE-NCC, zuständig für die Adressvergabe in Europa und Teilen Asiens, ihren letzten IPv4-Block und betonte erneut die Wichtigkeit des Umstieges auf IPv6. Während RIPE-NCC zur Umstellung auf IPv6 motiviert, steht man an anderer Stelle noch am Anfang. Betroffen sind insbesondere Internet-weite Scans, für die in IPv6 noch keine praktikablen Lösungen gefunden wurden, mit der schier endlosen Anzahl an möglichen Adressen umzugehen. Diese Arbeit baut auf der in der Bachelorarbeit entwickelten Adaption von ZMAP auf, um eine Internet-weite ICMPv6-Messung aller gerouteten /32-IPv6-Netzwerke durchzuführen. Im Gegensatz zu verwandten Arbeiten, die Echo-Replies auswerten, wird der Fokus auf die Interpretation der verschiedenen ICMPv6-Fehlermeldungstypen gelegt. Fehlermeldungen werden eingesetzt, um dem Sender mitzuteilen, warum ein Paket nicht zugestellt werden konnte. Bislang wurden diese bei Messungen ignoriert. Da sie jedoch eine potenzielle Informationsquelle zum Status des Zielnetzes darstellen, soll im Rahmen dieser Arbeit festgestellt werden, welche Typen sich in IPv6 im Einsatz befinden und was sich aus ihnen ableiten lässt. Die Ergebnisse dieser Arbeit zeigen, dass ICMPv6-Antworttypen sehr unterschiedliche Antwortzahlen liefern. Aus dem Antwortverhalten der Netzwerke konnten Cluster abgeleitet werden, welche es erlauben, aus Address-Unreachable-Antworten aktive Netzwerke abzuleiten. Dadurch wurde der Wert von Fehlermeldungen für Internet-weite Messungen bewiesen. Darüber hinaus wurden Cluster entdeckt, welche grobe Fehlkonfigurationen aufweisen. Eine Zuordnung der Netze zu den RIRs zeigt geografische Unterschiede. Während Netzwerke der LACNIC häufig von Rate-Limiting betroffen sind, sind Netzwerke der ARIN und APNIC stärker von Firewalling betroffen. RIPE-NCC hingegen weist neben Aliasing auch größere Anzahl an eher offen konfigurierten Netzwerken vor. Fehlkonfigurationen wie Routing-Loops verursachen zusätzliche Lasten im IPv6-Internet und sollten schnell behoben werden.

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Forschungsfrage 2

1.2 Methodik 2

1.3 Struktur dieser Arbeit 3

2 Grundlagen **5**

2.1 Grundlegende Funktionsweise des Internets 5

2.2 ICMPv6 6

2.3 IP-Adressraum 10

2.3.1 Regional-Internet-Registries 12

2.3.2 Warum wurde in IPv6-Netzwerken so viel Platz für Hosts gelassen? 13

2.3.3 Vergabe von Netzwerken 14

2.4 Internet-weite Messungen 16

2.4.1 Scan-Typen 16

2.4.2 Tools für Internet-weite Messungen 17

3 Messaufbau **19**

3.1 Ausgangspunkt der Messung 19

3.2 Zielnetzwerke 19

3.3 Messmethode 21

4 Anpassungen des Scan-Tools **23**

4.1 Interrupt-Funktionalität 23

4.2 Logging 26

5 Evaluierung der Messergebnisse **29**

5.1 Evaluierungskategorien 29

5.2 Vergleich Scan-Raten und Seeds 30

5.3 Aussortieren spezieller Netzwerkkonfigurationen 34
5.4 Antworten aus dem Zielnetzwerk 41
5.5 Antwortverhalten der Netzwerke 44
5.5.1 Herkunftsnetzwerke 45
5.5.2 Herkunft Antwort-Cluster 50
5.5.3 Zielnetzwerke 53
5.5.4 Herkunft Ziel-Cluster 55

6 Fazit 59

Abbildungsverzeichnis 61

Tabellenverzeichnis 63

Literatur 66

1 Einleitung

Internet-weite Scans erlauben uns einen Einblick in den momentanen Zustand des Internets. In IPv4 wird dadurch das Ausmaß an Schwachstellen überprüft sowie der Einsatz von Protokollen evaluiert - und bei jeder möglichen Adresse gemessen. Scans sind nicht limitiert bezüglich speziell dafür ausgelegter Hardware oder benötigtem Wissen. Dadurch wird im derzeitigen Internet ein hohes Maß an Transparenz und auch Resilienz erreicht, denn oben genannte Techniken werden in derselben Weise von Angreifern genutzt, um potenzielle Schwachstellen zu identifizieren. Dies birgt aber nicht unbedingt Nachteile, denn dadurch besteht ein sehr starkes globales Interesse daran, mit dem Internet verbundene Geräte stehts aktuell zu halten und den neuesten Standards anzupassen. Mittels IPv6 kann diese Information nicht mehr an jeder möglichen Adresse abgefragt werden, dazu reicht die momentan verfügbare Geschwindigkeit von der Internetinfrastruktur nicht aus. Deswegen wurden in den letzten Jahren mehrere Ansätze entwickelt, um die Anzahl der Ziele für Internet-weite Messungen zu reduzieren. Verwandte Arbeiten verfolgen im Wesentlichen zwei Richtungen, einerseits das Erstellen von Adresslisten, auch *Hit-Lists* genannt [1, 2], und andererseits das Ableiten möglicher Zieladressen aus gängigen Adressvergabemustern, was unter den Begriff *Target-Generation* Algorithmen fällt [3, 4], oder einer Mischung aus beidem [5].

Hit-Lists dienen als Grundlage für weitere Messungen, bergen jedoch die Gefahr, Vollständigkeit vorzutäuschen. Es kann nicht garantiert werden, dass die gesammelten Adressen wirklich den momentanen Zustand des gesamten IPv6-Internets repräsentieren. *Target-Generation* Algorithmen bauen auf Adressmustern auf, um weitere Adressen daraus abzuleiten. Dies bietet unter anderem die Möglichkeit weitere aktive Hosts in einem Netzwerk zu finden, sollten bereits aktive Adressen aus diesem bekannt sein. Deren Effektivität wurde jedoch anhand aktueller Adressmuster evaluiert und könnte sich daher in Zukunft ändern.

In der Bachelorarbeit wurde ein Scan-Tool entwickelt, um IPv6-weite Messungen durchzuführen [6]. Der Grundgedanke dabei war, IPv6-Messungen in mehrere Schritte einzuteilen. Erstens sollen anhand einer Liste von IPv6-Netzwerken aktive Blöcke in diesen erkannt werden, dazu zählen Subnetze oder kleinere Netzwerke, welche im größeren allokiert wurden. Im zweiten Schritt kann auf bereits vorhandene Tools zurückgegriffen werden, um in aktiven Adressbereichen Hosts zu finden. Auch für das Erkennen von aktiven Subnetzen oder kleineren Netzwerken muss die Anzahl der Pakete stark reduziert werden, da zum Beispiel bei einem /32 Netzwerk 2^{32} mögliche Subnetze vorhanden sind, welche wiederum jeweils 2^{64} mögliche Hosts aufweisen können. Das Scan-Tool beschränkt sich pro Subnetz auf die

::1 Adresse. Um die Funktionalität des Scantools zu zeigen, wurde in der Bachelorarbeit in einem ersten Scan des österreichischen Adressraums nur auf positive Antworten zurückgegriffen, um aktive Netzwerke zu finden. Bei so vielen Möglichkeiten ist die Chance beim Scannen des ::1 Hosts eine positive Antwort zu erhalten, jedoch sehr gering und nur möglich, da Router-Adressen oder DHCP-Konfigurationen häufig ::1 als *Host-Identifier* verwenden [7]. In dieser Arbeit soll daher evaluiert werden, welche ICMPv6-Antworttypen neben Echo-Replies bei einem Scan aller gerouteten IPv6-Netzwerke empfangen werden und ob dadurch Rückschlüsse auf das Netzwerk möglich sind.

1.1 Forschungsfrage

Der Vergleich von bisherigen Ansätzen zeigt, dass diese zwar Internet-weite IPv6-Messungen erlauben, jedoch nicht das gleiche Maß an Transparenz wie in IPv4 aufweisen. Bano et al. [8] zeigten erstmals den Wert von ICMPv6-Fehlermeldungen bei Messungen in IPv4, verfolgten diesen Ansatz jedoch nicht in IPv6. Im Rahmen dieser Arbeit werden daher folgende zwei Fragen gestellt.

1. Welche ICMPv6-Antworttypen werden von IPv6-Netzwerken gesendet und was sagen diese über den Zustand der Zielnetzwerke aus?

2. Wie stark ist der Einfluss von unterschiedlichen Scanraten, bzw. Rate-Limiting auf die einzelnen Fehlermeldungstypen und wie kann bei dieser Scanmethode Aliasing von einzelnen Netzwerken erkannt werden?

Bei der zweiten Frage soll geklärt werden, welche Scanparameter zu den besten Ergebnissen führen, beziehungsweise wie sichergestellt werden kann, dass einzelne Netzwerke mit speziellen Konfigurationen wie Aliasing die Ergebnisse nicht verfälschen. Bei Aliasing wird auf jedes Paket eine positive Antwort zurückgeschickt, egal ob es sich um ein aktives oder nicht aktives Ziel handelt. Dies reduziert die Aussagekraft von positiven Antworten.

1.2 Methodik

Im Rahmen dieser Arbeit wird eine Internet-weite Messung aller gerouteten /32 Netzwerke durchgeführt. In dieser Messung werden nicht nur positive Antworten, sondern auch ICMPv6-Fehlernachrichten aufgezeichnet. Dabei soll festgestellt werden, welche Fehlernachrichten häufiger eingesetzt werden. Daraus wird der Wert der einzelnen Fehlernachrichtentypen für Internet-weite Scans abgeleitet. Als Protokoll für die Messung wurde ICMPv6 festgelegt, da jeder mit dem IPv6-Internet verbundene Host ICMPv6 unterstützen muss [9]. Die Schwerpunkte der Messung befassen sich mit folgenden Themen:

- **ICMPv6-Antworten** - Es soll ausgewertet werden, wie viele Antworten welchen Typs zurückgeschickt werden, und deren Aussagekraft erhoben werden. Das Antwortverhalten der Netzwerke soll analysiert und untersucht

werden, ob sich Zusammenhänge finden lassen.

- **Rate-Limiting** - Welcher Einfluss hat Rate-Limiting auf die einzelnen Typen an Fehlernachrichten und gehen damit Informationen über das Zielnetzwerk verloren?

- **Aliasing** - Es soll sichergestellt werden, dass Netzwerke mit Aliasing erkannt und aus den Ergebnissen aussortiert werden können. Es soll überprüft werden, ob bereits vorhandene Verfahren dafür eingesetzt werden können.

- **Anforderungen an den Scanner** - Solche Messungen in IPv6 können oftmals mehrere Wochen dauern. Es soll ermöglicht werden, die Adaption von ZMAP zu unterbrechen, um zum Beispiel Serverwartungen durchzuführen und den Scan anschließend wieder fortzusetzen. Solch umfangreiche Scans produzieren eine Menge an Daten, es soll über verschiedene Logging-Lösungen diskutiert werden und herausgefunden werden, welche sich am besten für einen Internet-weiten IPv6-Scan eignet.

Die genannten Schwerpunkte werden jedoch nicht in der Auflistung befindlichen Reihenfolge in dieser Arbeit abgehandelt. Folgend wird beschrieben in welche Schwerpunkte in welchen Kapiteln behandelt werden.

1.3 Struktur dieser Arbeit

Diese Arbeit ist in mehrere Abschnitte gegliedert. In Kapitel 1 wurden die Herausforderungen für Internet-weite IPv6-Scans erläutert, bisherige Lösungsansätze miteinander verglichen und die Rahmenbedingungen sowie die Ziele dieser IPv6-Messung definiert.

In Kapitel 2 werden die Grundlagen für diese Messung besprochen. Dabei werden die Protokolle beschrieben, auf denen die Messung aufbaut, die Struktur des IPv4- und IPv6-Adressraums verglichen und Internet-weite Messungen in IPv6 näher beleuchtet.

Kapitel 3 definiert die Rahmenbedingungen der Messung. Dabei wird das Scan-Setup beschrieben. Dazu gehören die Beschreibung der Ausgangspunkt der Messung, welche Methode zum Scannen benutzt wird und was die Input-Netzwerke sind.

Kapitel 4 beschäftigt sich mit der Umsetzung der Messung. Es wird darauf eingegangen, welche Änderungen für diese Messung an ZMAP vorgenommen wurden und welche Optimierungen gemacht wurden.

In Kapitel 5 werden die Evalierungsmethoden und Ergebnisse der Messung präsentiert. Die Evaluierung beinhaltet die Definition von Evaluierungskategorien. Es wird gezeigt wie sich die einzelnen Antworttypen in diesen Kategorien verhalten und die Einflüsse von Rate Limiting auf die Ergebnisse beobachtet. Darüber hinaus wird das Aussortieren von besonderen Netzwerkkonfigurationen ermöglicht. Nach dem Aussortieren wird auf den Beitrag der einzelnen Antworttypen zu der Anzahl an Antworten, antwortenden Adressen, Netzwerke und Subnetze eingegangen. Anschließend wird ausgewertet, wie viele Antworten aus dem Zielnetzwerk stammen und eine Methode vorgestellt, um die Nähe der antwortenden Adresse zur Zieladresse festzustellen. Abschließend werden die Netzwerke je nach Antwortverhalten in Gruppen eingeteilt. Es wird gezeigt, wodurch die einzelnen Gruppen entstehen, sowie die geografischen Unterschiede in den Gruppen ausgewertet.

Kapitel 6 reflektiert die Ergebnisse dieser Arbeit, zeigt welchen Einfluss diese auf das IPv6-Internet haben und gibt einen Ausblick für zukünftige Messungen, die sich aus dieser Arbeit ergeben.

2 Grundlagen

In diesem Abschnitt werden die Technologien, auf denen diese Arbeit beruht, näher beleuchtet. Internet-weite Messungen mussten sich an die rasante Entwicklung des Internets anpassen, doch der Ursprung der im Einsatz befindlichen Protokolle, wie ICMP und BGP, liegt zu Beginn des Internets. Diese Protokolle werden in nachfolgend referenzierten Request for Comments (RFC) definiert, welche Standards im Internet festlegen.

2.1 Grundlegende Funktionsweise des Internets

Das in RFC 792 definierte *Internet-Control-Message-Protocol (ICMP)* wurde im September 1981, im gleichen Monat wie das *Internet-Protocol (IP)*, veröffentlicht [10]. Der starke Zusammenhang dieser zwei Protokolle beruht jedoch nicht nur auf deren gemeinsamen Veröffentlichung, sondern auch auf deren Funktionsweise. Beide Protokolle wurden geschaffen, um die Kommunikation mehrerer miteinander verbundenen Netzwerke zu ermöglichen, welche damals unter dem heute eher unbekannten Begriff Catenets ins Leben gerufen wurden. Die Kommunikation wurde erstmals mittels Paketvermittlung umgesetzt, was damals eine große Neuerung war, denn die erste Verbindung zwischen Computern wurde noch per Telefonkabel und Leitungsvermittlung verwirklicht [11]. Dies wäre mit der heutigen Anzahl an miteinander verbundenen Geräten undenkbar, denn für jede aktive Verbindung müsste eine komplette Leitung für diesen Kommunikationskanal reserviert werden. Bei der Paketvermittlung kommunizieren mehrere Teilnehmer über die gleiche Leitung, was es ermöglicht, Bandbreite zu sparen. In Grafik 2.1 wurde dieser Unterschied visualisiert.
Die Visualisierung soll verdeutlichen, dass bei der Leitungsvermittlung jeder Kommunikationskanal eine direkte Verbindung darstellt, während sich bei der Paketvermittlung mehrere Pakete eine *Leitung* teilen. Es muss daher pro Paket die Entscheidung getroffen werden, wohin dieses geleitet werden soll. Typischerweise enthält ein Paket dafür Herkunfts- und Zieladresse, Länge, Sequenznummer, Art des Paketes und einen Datenteil [12]. Diese Felder finden sich auch im Protokollaufbau von IP wieder, welches das Format für den Paketaustausch im Internet definiert. Größere Nachrichten werden bei der Paketvermittlung einfach in mehrere Pakete aufgeteilt. Die einzelnen Pakete passieren dabei unabhängig voneinander die einzelnen Vermittlungsstellen, welche anhand der Zieladresse entscheiden, wohin das Paket weitergeleitet werden soll. Die Sequenznummer teilt dem Empfänger mit, in welcher Reihenfolge die Pakete zu interpretieren sind, da diese nicht zwingend in der gleichen Reihenfolge ankommen müssen, in der sie abgeschickt

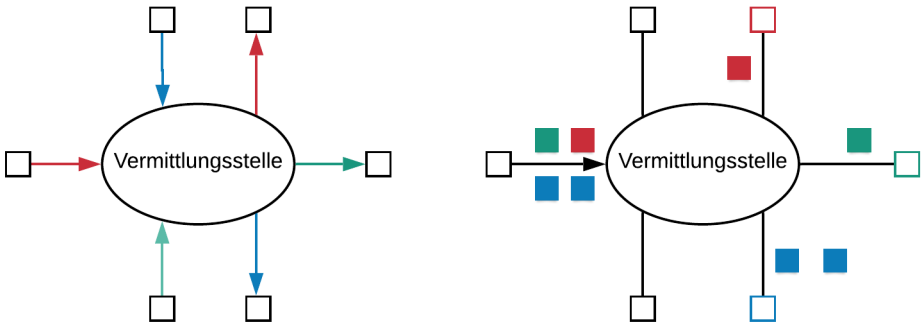


Abbildung 2.1: Vermittlungstypen (links: Leitungsvermittlung - rechts: Paketvermittlung)

wurden. Der Einsatz von Paketvermittlung im Internet hat jedoch auch Folgen für Internet-weite Messungen. Bei solchen wird an jede mögliche Zieladresse ein oder mehrere Pakete geschickt. Je nach Topologie der Vermittlungsstellen können an Knotenpunkten sehr hohe Lasten entstehen. Dies kann jedoch nicht nur die Kommunikation mit dem Ziel beeinflussen, sondern im schlimmsten Fall alle Pakete, die diese Vermittlungsstelle passieren. Es sollte daher bei Internet-weiten Messungen darauf geachtet werden, die Ziele so auszuwählen, dass die Auslastung minimiert wird.
Der erste und wohl bekannteste Zusammenschluss an Netzwerken ist das ARPANET, welches die technische Grundlage für das größte Catenet bildet, dem heutigen Internet. Die Geräte, welche die Schnittstellen zu anderen Netzwerken darstellen, werden *Gateways* genannt. Damit diese wissen, welche Pakete an wen weitergeleitet werden sollen, müssen diese ebenfalls miteinander kommunizieren. Dafür wurde das *Gateway-to-Gateway* Protokoll (GGP) definiert, welches regelt, wie Distanzinformationen untereinander ausgetauscht werden. Es ist heute noch in seiner jetzigen Form, dem *Border-Gateway-Protocol* (BGP), im Einsatz [13]. Da die Kommunikation mehrerer Netzwerke aber nicht immer reibungslos ablief, wurde ICMP definiert. Es ermöglicht die Prüfung, ob eine Verbindung mit dem Zielnetzwerk hergestellt werden kann und wenn dies nicht der Fall ist, dem Sender mitzuteilen, warum sein Paket nicht zugestellt werden konnte. Es muss jedoch keine Antwort zurückgeschickt werden. Für Zuverlässigkeit wird erst die über IP liegende Schicht verantwortlich gemacht. Allerdings wurde festgelegt, dass jedes IP-Modul auch ICMP unterstützen muss [10]. Während bei BGP ausschließlich Gateways miteinander kommunizieren, kommunizieren bei ICMP entweder zwei Hosts direkt miteinander oder ein Gateway mit einem Host. Als Host wird ein Rechner innerhalb eines Netzwerkes gesehen.

2.2 ICMPv6

In IPv6 ist es schwierig auf aktive Adressen in einem Netzwerk zu stoßen. Es wurde bei Messungen jedoch festgestellt, dass ICMPv6-Echo-Requests zu mehr Antworten führen als im Vergleich zu anderen Protokollen wie TCP oder UDP

[1]. Im Gegensatz zu anderen Protokollen muss jedes mit dem Internet verbundene Gerät auch ICMP(v6) unterstützen. ICMP(v6) Nachrichten werden benutzt, um den Verbindungstatus mit einem Ziel zu überprüfen. Die am meisten verwendete Funktion von ICMP(v6) sind das Verschicken von Echo-Request Paketen. Sollte das Ziel den Echo- Request erhalten wird standardmäßig mit einem Echo Reply geantwortet. Dadurch kann sowohl auf ein aktives Ziel als auch auf eine funktionierende Verbindung mit dem Ziel geschlossen werden. An Netzwerkgrenzen werden Echo-Requests oftmals gefiltert, da nicht von außerhalb des Netzwerkes nicht abgefragt werden können soll, welche aktiven Hosts in dem Netzwerk vorhanden sind. Es kann auch passieren, dass der Echo-Request am Weg zum Ziel nicht weitergeleitet wird, weil zum Beispiel keine Route für das Zielnetzwerk vorhanden ist. Gateways haben die Möglichkeit den Sender zu informieren, sollte das Paket nicht weitergeleitet oder gefiltert worden sein, indem eine Error-Message zurückgeschickt wird. Je nach Grund, warum das Paket nicht zugestellt werden konnte, unterscheiden sich sowohl in ICMP als auch ICMPv6 das Typ- und das Code-Feld. Die Vorgaben, wann welcher Code benutzt werden sollte, wurden in RFC 792 (ICMP) und RFC4443 (ICMPv6) definiert [10, 9]. Vor allem die für ICMP beschriebenen Typen wurden mehrmals überarbeitet und teils für veraltet erklärt. Die Anzahl an Typen hat sich mit ICMPv6 auf ein Minimum reduziert. Da der Fokus dieser Arbeit auf IPv6 liegt, beschränkt sich die folgende Beschreibung auf ICMPv6-Typen, die im Rahmen der Messung als Antwort erwartet werden. In Tabelle 2.1 werden diese zusammengefasst und mit Abkürzungen versehen, welche im weiteren Verlauf dieser Arbeit verwendet werden.

Tabelle 2.1: ICMPv6 Typen [9]

Typ	Bezeichnung	Code	Verwendete Abkürzung
1	Destinatio-Unreachable	0	No-Route
		1	Admin-Prohib
		2	Scope
		3	Address-Unreach
		4	No Port
		5	Ingress / Egress
		6	Reject-Route
2	Packet-Too-Big	0	Too Big
3	Time-Exceeded	0	Time-Exceeded
4	Parameter-Problem	0	Parameter
129	Echo-Reply	0	Echo-Reply

Da der Typ Destination-Unreachable sechs verschiedene Sub-Codes aufweist und diese unterschiedlichen Aussagen über das Ziel liefern, wurde für jeden dieser eine eigene Abkürzung erstellt. Typen 2 und 4 wurden zwar mit einer

Abkürzung versehen, sie sollten allerdings nicht im Rahmen dieser Messung zurückgeschickt werden. Typ 2 Packet-Too-Big wird verwendet, wenn das Paket größer als die *Maximum-Transmission-Unit* (MTU) des Links zum nächsten Hop ist. Tools wie ZMAP erstellen jedoch möglichst kleine Pakete, um die Bandbreite nicht unnötig auszulasten. Typ 4 Parameter-Problem weist drei Sub-Codes auf, welche auftreten, wenn die Verarbeitung des Paket-Headers ein Problem aufweist. Mit diesem wird im Rahmen dieser Arbeit ebenfalls nicht gerechnet, da RFC konforme Echo-Requests ausgeschickt werden. Je nach Typ der Error-Message unterscheidet sich meist auch die Herkunft. In Grafik 2.2 wird dies visualisiert.

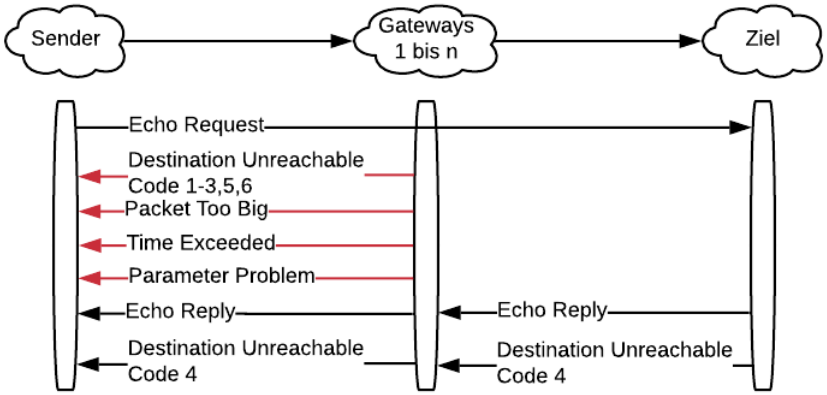


Abbildung 2.2: Ursprung Antworttypen

Error-Messages dienen dazu, den Sender über fehlerhafte Paketzustellung zu informieren. Während Echo-Replies meist direkt vom Ziel stammen, werden Destination-Unreachable-Codes 1-3,5 und 6 von Gateways benutzt [14]. In Grafik 2.2 wird ersichtlich, dass nur ein Error-Message-Typ direkt vom Ziel-Host erwartet wird. Der Rest sollte gemäß RFC4443 von Gateways am Weg zum Ziel stammen. Neben Herkunft und Aussage über das Ziel gibt es noch ein anderes Merkmal, das Error-Messages von Echo-Replies unterscheidet, nämlich der Aufbau des Headers, welcher in Grafik 2.3 dargestellt wird.

Neben Typ- und Code-Feld verfügt der ICMPv6-Header über eine 16 Bit Prüfsumme. Die Prüfsumme wird unter anderem von ZMAP benutzt, um einen eindeutigen Wert für die Echo-Requests zu generieren. Der Wert wird durch die Empfangseinheit evaluiert, um zu überprüfen, ob es sich um eine valide Antwort handelt. Dadurch kann ZMAP *stateless* operieren, das heißt, Sende- und Empfangseinheit arbeiten unabhängig voneinander und es können höhere Scanraten erreicht werden [15]. Wie sich Grafik 2.3 entnehmen lässt, enthalten Error-Messages nach der Prüfsumme das ursprüngliche Anfragepaket. Dadurch lässt sich unter anderem feststellen, welchen Hop-Zähler das Anfragepaket hatte, nachdem es das Ziel erreichte. Der Aufbau des Headers ist für alle Error-Messages gleich. Folgend werden die Unterschiede zwischen einzelnen ICMPv6-Typen und -Codes näher betrachtet. Die Beschreibung fokussiert sich auf die Vorgaben in RFC4443 [9].

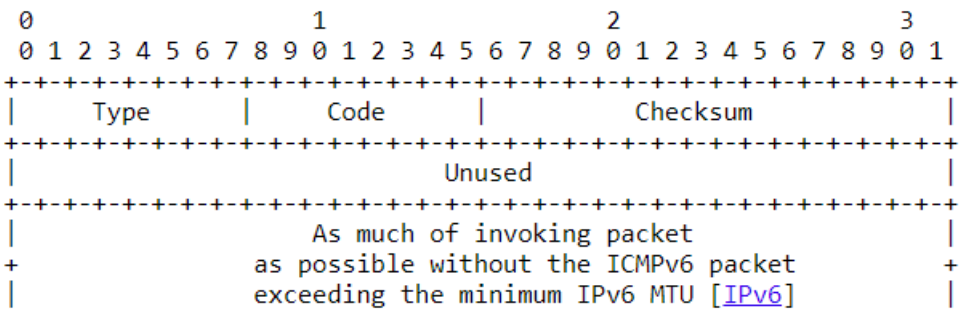


Abbildung 2.3: ICMPv6-Error-Message-Header [9]

- Destination-Unreachable - Teilt sich in sechs verschiedene Untertypen, auch Codes genannt, und beinhaltet damit die meisten Untergruppen aller Error-Message-Typen.
  - No-Route - wird benutzt, um dem Sender mitzuteilen, dass für dieses Zeil kein Eintrag in der Routing-Tabelle des Gateways vorhanden ist.
  - Admin-Prohib - kann gesetzt werden, um dem Sender mitzuteilen, dass ihm nicht erlaubt wird, mit dem Ziel zu kommunizieren.
  - Scope - ist ein Spezialfall, indem zum Beispiel eine linklokale Adresse mit einer globalen zu kommunizieren versucht.
  - Address-Unreach - kann verwendet werden, wenn kein anderer Grund zutreffen sollte. Das kann zum Beispiel vorkommen, wenn die zur Ziel-Adresse gehörige Linkadresse nicht aufgelöst werden kann. Address-Unreach muss jedoch verwendet werden, wenn ein Paket von einem Point-to-Point-Link empfangen wird und die Antwort an eine Adresse im Subnetz des Point-to-Point-Link zurückgeschickt werden müsste, also an den vorigen Router zurückgeleitet werden würde.
  - No-Port - Ist der einzige Code, der auch vom Ziel-Host selbst benutzt werden kann. Dadurch wird dem Sender mitgeteilt, dass für das Transportprotokoll kein *Listener* vorhanden ist.
  - Ingress / Egress - Sollten gewisse Eingangs- oder Ausgangsregeln verhindern, dass das Paket zugestellt wird, kann Code 5 genutzt werden.
  - Reject-Route - Router erlauben das Konfigurieren sogenannter *Reject*-Routen. Alle Pakete mit Zieladressen, welche über Reject-Routen weitergeleitet werden sollten, werden nicht weitergeleitet. Code 6 kann benutzt werden, um dem Sender mitzuteilen, dass das Paket aus diesem Grunde nicht weitergeleitet wurde. Sowohl Code 5 als auch Code 6 sind Erweiterungen zu Code 2, um dem Sender spezifischere Informationen im Falle eines Filters zu liefern.

- Time Exceeded - Dieser Typ besteht ebenfalls aus zwei verschiedenen Codes. Code 0 wurde bereits erwähnt und tritt auf, wenn das Hop-Limit eines Paketes auf null gesetzt wird. Code 1 kann auch vom Ziel selbst stammen. Dieser wird verwendet, wenn die einzelnen Fragmente eines großen Paketes nicht in einer gewissen Zeitspanne eintreffen und diese daher nicht wiederhergestellt werden können. Da Code 1 ohne Fragmentierung der Anfragen jedoch nicht auftreten kann, wurde dieser in Grafik 2.2 nicht gezeigt.

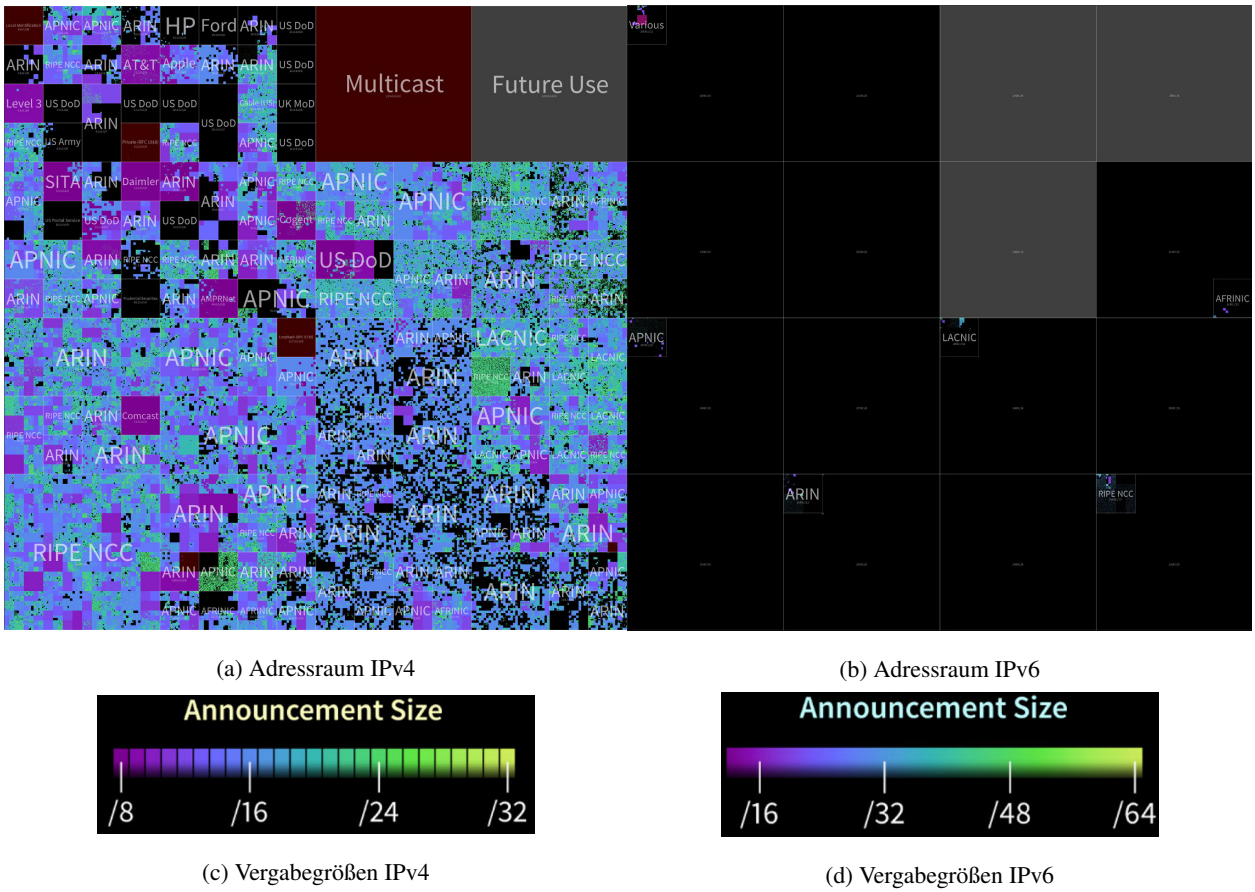
2.3 IP-Adressraum

In seiner ursprünglichen Fassung RFC 675 beinhaltete das *Transmission-Control-Program* die Funktion von IP. Es gab also keine Trennung zwischen TCP und IP. Die Kommunikation sollte über Sockets erfolgen, welche aus einer Netzwerknummer und einer TCP-Portnummer bestehen [16]. Erst mit der Trennung von TCP und IP wurde der Grundstein für den modularen Aufbau des Internets gebildet und es entstand IPv4, wie wir es in seiner heutigen Form kennen. IP bildet die Grundlage für Protokolle wie TCP, welches für das Ansprechen von Ports verantwortlich ist. Die Wahl des Zieles erfolgt anhand einer Zieladresse mit einer fixen Länge von 32 Bit. Mittels IPv4 können durch die fix vorgegebene Länge somit 2^{32} verschiedene Adressen angesprochen werden [17]. Werden mehr Adressen benötigt, kann die Länge jedoch nicht einfach angehoben werden. Durch den rasanten Zuwachs an verbundenen Geräten wurde eine neue Version definiert. Für IPv6 wurde eine fixe Länge von 128 Bit festgelegt [18]. Die Vergabe der Adressen erfolgte von Beginn an von einer zentralen Stelle. Diese Aufgabe sollte zu Zeiten des ARPANETS von einer einzelnen Person geregelt werden, nämlich von Jon Postel. Dieser legte mit seiner Arbeit den Grundstein für die *Internet-Assigned-Numbers-Authority*, welche als gemeinnützige Organisation auch heute noch für die Adressvergabe im Internet zuständig ist [19]. In dieser Arbeit wird für Internet-weite Messungen folgendes daraus abgeleitet:

- Die Adressen im Internet werden von einer zentralen Stelle vergeben.
- Der Adressraum für IPv4 und IPv6 verfügt über eine hierarchische Struktur. Diese wird von der IANA geprägt.
- Die Anzahl der möglichen Adressen in einem IPv6-Netzwerk übersteigt jegliches Vorstellungsvermögen.
- Messungen sollten sich nicht auf den gesamten, sondern nur auf den bereits vergebenen Adressbereich beschränken, um unnötige Auslastungen zu vermeiden

Der Schwerpunkt wird nun auf den Vergleich zwischen dem IPv4- und dem IPv6-Adressraum gelegt. Um die Struktur beider Adressräume besser verstehen zu können, wird eine Visualisierung der Universität Oregon in Grafik 2.4 gezeigt.

In Grafik 2.4 wird auf der linken Seite der IPv4- und auf der rechten Seite der IPv6-Adressraum dargestellt, wobei beide im Original ursprünglich 4096x4096 Pixel beinhalten. Dies ergibt in Summe 16,7 Millionen einzelne Pixel.



(a) Adressraum IPv4

(b) Adressraum IPv6

(c) Vergabegrößen IPv4

(d) Vergabegrößen IPv6

Abbildung 2.4: Visualusierung des Adressraums durch die Universität Oregon im Jänner 2019 [20].

Was sofort auffällt ist, dass, während der IPv4-Adressraum fast zur Gänze genutzt wird, IPv6 vergleichsmäßig leer wirkt. Dadurch lassen sich die Farben der Netzwerke für IPv6 auch sehr schwer erkennen. Nachfolgend werden beide näher betrachtet. Für IPv4 repräsentiert ein einzelnes Pixel 256 Adressen, was umgerechnet einem /24 Netzwerk entspricht. Für IPv6 wäre das Äquivalent 2^{104} Adressen pro Pixel. Nachdem jedoch nicht der gesamte IPv6-Adressraum benutzt wird, beschränkt sich die Grafik auf 2000::/4, was ein Sechzehntel des Adressraums ist, aus welchem aktuell Adressen vergeben werden [20]. Das bedeutet, dass ein Pixel 2^{100} Adressen entspricht, umgerechnet also einem /28-IPv6-Netzwerk. Die Eingrenzung auf 2000::/4 wurde getroffen, um genügend weitere /4 Adressbereiche in Reserve zu haben, falls aktuelle Vergabemuster sich erneut als zu verschwenderisch erweisen sollten. Dies ist auch nicht unwahrscheinlich, denn anhand der Farben in Grafik 2.4 erkennt man aktuelle Vergabegrößen. Das Spektrum startet bei Violett mit den größten Netzwerken und endet bei Gelb mit den kleinsten, wobei schwarz für den nicht verwendeten Adressbereich steht. Die Skala bewegt sich für IPv4 im Bereich von /8 bis /32 und für IPv6 von /16 bis /64. Sowohl in IPv4 als auch in IPv6 wurden riesige Adressbereiche vergeben. In IPv4 stechen zunächst der Multicast und der Block

für *Future-Use* heraus. Dies sind zwei der 16 /4-Blöcke in denen sich IPv4 einteilen lässt. Ein /4-Block lasst sich wieder in 16 weitere /8-Blöcke aufteilen. Und hier liegt die nächste deutlich merkbare Vergabegrenze. In der Grafik lassen sich zum Beispiel /8-Blöcke mit der Beschriftung Ford, HP oder US-Army erkennen. In den Anfangszeiten des Internets wurden solche riesigen Blöcke noch an einzelne Unternehmen und Organisationen vergeben. Damals erfolgte die Einteilung in die Adressklassen A, B und C, wobei /8 Klasse-A entspricht und Platz für 16,7 Millionen Hosts bietet und /24 Klasse C mit 256 möglichen Hosts entspricht [21]. In IPv4 lässt sich die Vergabe riesiger Blöcke somit auf die Adressklassen zurückführen. Diese wurden durch die CIDR-Notation ersetzt, welche mit dem typischen Format /xx variable Netzwerkgrößen ermöglichte. Weitere Blöcke, die sich deutlich hervorheben, gehören zu ARIN, APNIC, LACNIC, RIPE NCC und AFRINIC. Diese repräsentieren die fünf *Regional-Internet-Registries* (RIRs), welche Adressblöcke von der IANA bekommen und diese weitervergeben.

2.3.1 Regional-Internet-Registries

Weltweit gibt es fünf RIRs, welche hierarchisch der IANA untergeordnet sind und für die einheitliche Adressvergabe in ihrer Region zuständig sind. Pro RIR wird eine WHOIS-Datenbank geführt, welche die Allokationen der RIRs dokumentiert. Die Weitergabe erfolgt typischerweise an Internet-Service-Provider (ISPs) oder in Ausnahmefällen direkt an einzelne Organisationen. Die Allokation an die RIRs durch die IANA werden von der IANA öffentlich zu Verfügung gestellt [22]. IP-Adressen können anhand ihrer Zugehörigkeit zu einem RIR geografisch zugeordnet werden. In Grafik 2.5 werden die Gebiete der RIRs dargestellt. RIPE-NCC verfügt über das geografisch größte Gebiet. Dieses umfasst Europa, den Mittleren Osten und Teile Asiens. Südlich davon befindet sich das Gebiet der AFRINIC, das wie im Namen enthalten, den Kontinent Afrika abdeckt. APNIC steht für das Asia-Pacific-Network-Information-Centre und ist für Südostasien und den Pazifikraum zuständig. Das Gebiet der ARIN beinhaltet die Staaten Kanada und USA. LACNIC ist für Mittel- und Südamerika zuständig. Die Grafik 2.5 zeigt, dass einige der Gebiete mehr als einen Kontinent umfassen. Anhand der RIRs kann nur eine ungefähre geografische Einteilung erfolgen. IP-Adressen können allein durch die Zuordnung zu RIRs nicht eindeutig einem Kontinent zugeordnet werden.

Die RIRs sind sowohl für IPv4- als auch für IPv6-Adressvergaben zuständig. Jeder RIR wurde jeweils ein /12-Block zugeordnet. Laut RIPE-Regelwerk befinden sich die Allokationen an *Local-Internet-Registries* (LIRs), auch ISPs genannt, im Rahmen zwischen /29 und /32 und nur mit Angabe spezieller Begründungen können größere Netzwerke vergeben werden [24]. Die Vergabe an Endkunden erfolgte laut RFC6177 auf /48 Basis, wurde jedoch überarbeitet und nun verkleinert auf /56 [25]. Es lässt sich also in IPv6 erneut feststellen, dass ursprüngliche Vergabegrößen angepasst werden müssen und mit der zunehmenden Nutzung des Adressraums wird sich zeigen, ob man wirklich aus den Fehlern von IPv4 gelernt hat. Die sehr großzügig ausgelegten Vergabegrößen in IPv6 lassen sich jedoch auf folgendes zurückführen.

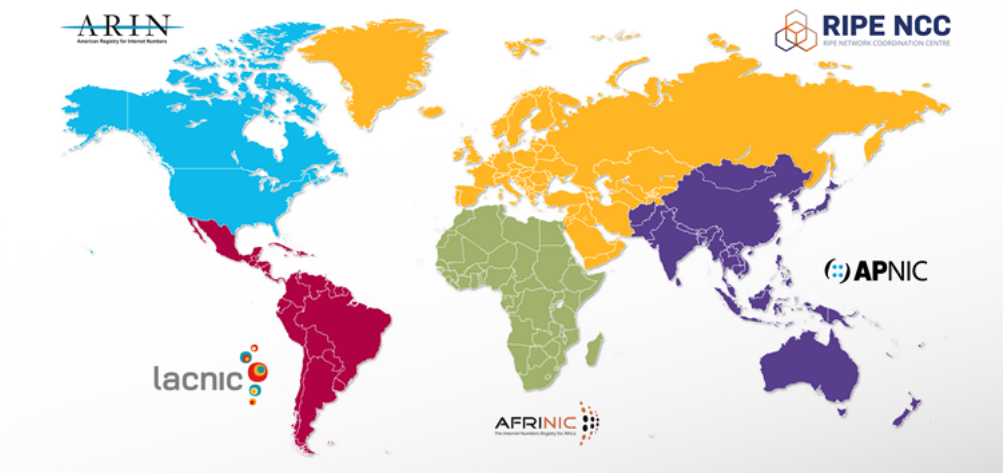


Abbildung 2.5: Regional-Internet-Registry Regionen[23]

2.3.2 Warum wurde in IPv6-Netzwerken so viel Platz für Hosts gelassen?

In RFC2373 *IP-Version-6-Addressing-Architecture*, veröffentlicht im Juli 1998, also ein halbes Jahr vor RFC2460, in welchem das IPv6-Protokoll definiert wird, wurden bereits die Adresstypen und deren Format angegeben. Diese gliedern sich in Multicast, Anycast und Unicast. Sowohl Multicast als auch Unicast verfügen über ihren eigenen Adressraum, wobei man beim IPv6-Internet typischerweise vom Unicast-Adressraum spricht. Für diesen wurde in RFC2373 der Präfix 001 (ein Achtel des gesamten Adressraums) definiert und die IANA verwendet, wie in diesem Kapitel bereits erwähnt, den 0010 (1/16) Bereich für die Adressvergabe. In dem gleichem RFC wurde auch das Format für Unicast-Adressen visualisiert, welches in Grafik 2.6 gezeigt wird [26]. Wie sich in Grafik 2.6 erkennen lässt, war

```
| 3|   13  | 8 |    24    |   16   |          64 bits                |
+--+-------+---+----------+--------+---------------------------------+
|FP| TLA   |RES|  NLA     |  SLA   |          Interface ID           |
|  | ID    |   |  ID      |  ID    |                                 |
+--+-------+---+----------+--------+---------------------------------+

Where

   001          Format Prefix (3 bit) for Aggregatable Global
                Unicast Addresses
   TLA ID       Top-Level Aggregation Identifier
   RES          Reserved for future use
   NLA ID       Next-Level Aggregation Identifier
   SLA ID       Site-Level Aggregation Identifier
   INTERFACE ID Interface Identifier
```

Abbildung 2.6: Aggregierbare Global-Unicast-Adressen [26, p.7]

der Aufbau von Unicast-Adressen sehr kompliziert. Bereits 2003 wurde dieses Format in RFC3587 überarbeitet und vereinfacht. Als Gegenüberstellung wird in Grafik 2.7 das neue Format abgebildet.

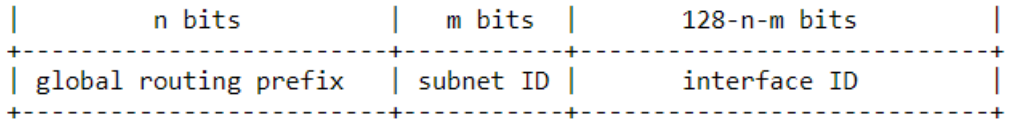


Abbildung 2.7: Global-Unicast-Adressformat Neu [26, p.7]

Wie in Grafik 2.7 zu sehen ist, wurden die ersten 48 Bit zu einem *global routing prefix* mit variabler Länge zusammengefasst. Die *subnet ID* spiegelt die *SLA ID* aus Grafik 2.6 wieder, behält seine Funktion als Subnetzbits bei und verfügt nun ebenfalls über variable Länge [27]. Was gleich geblieben ist, ist der *Interface-Identifier*. Schon in RFC2373 wurde definiert, dass dieser das IEEE EUI64 Format unterstützen soll [26]. Das Format ermöglicht das Einbinden einer 48 Bit langen MAC-Adresse in den *Interface-Identifier*. Durch EUI-64 kann jede globale Unicast-Adresse anhand der eingebundenen MAC-Adresse eindeutig identifiziert werden. Aus Privacy-Gründen hat es sich daher nie zur Gänze durchgesetzt [2].

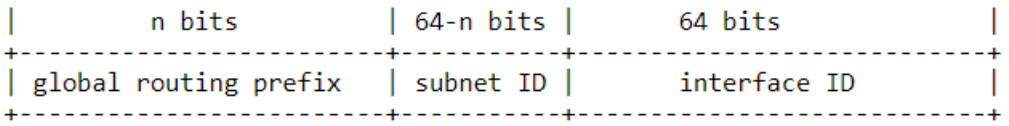


Abbildung 2.8: Global-Unicast-Adressformat EUI-64 [26, p.7]

In Grafik 2.8 wird das an EUI64 angepasste Format gezeigt. Es wurde in RFC3587 erneut definiert, dass globale Unicast-Adressen, welche nicht mit dem 000 Präfix beginnen, über einen dem EUI64 Format entsprechenden *Interface-Identifier* verfügen müssen [27]. Daraus ergibt sich, dass für den Netzwerkpräfix nur noch 128-64 (Interface-ID) - m (Subnetz-ID) Bits übrig bleiben. Die Anzahl an möglichen Netzwerken wurde dadurch stark reduziert und die Anzahl an möglichen Hosts in einem Netzwerk ist dementsprechend hoch.

2.3.3 Vergabe von Netzwerken

Die Vergabe von Netzwerken erfolgt im ersten Schritt durch die IANA, aber im zweiten, viel wesentlicheren, durch die einzelnen RIRs. Diese regeln die Vergabe von Adressblöcken an ISPs. Als Maßzahl, wie weit ein ISP einen Adressbereich weiterallokiert hat, wurde in RFC1715 zuerst das H-Ratio und anschließend durch geringfügige Adaption in RFC3194 das HD-Ratio definiert [28, 29].

$$\mathrm{HD} = \frac{\log(\text{Nr allokiert})}{\log(\text{Nr allokierbar})} \tag{2.1}$$

Die Logarithmusfunktion wird angewandt um große Zahlen besser zu veranschaulichen. Der HD-Wert gibt das Verhältnis von bereits verteilten Adressen und der zu Verfügung stehenden Gesamtmenge an Adressen an. Der resultierende Wert liegt im Bereich 0 <= x <= 1. Schon beim H-Ratio wurde ein Schwellwert von 80%, also 0.8, festgelegt, ab dem eine Allokation als aufgebraucht gilt [28]. In folgendem Szenario wird angenommen, dass der IPv6-Adressraum keine hierarchische Struktur besitzt und ein HD-Ratio auf diesen angewandt wird. Dabei würden die Netzwerke von einer zentralen Stelle direkt an die Endkunden vergeben werden. Die Ergebnisse werden in Tabelle 2.2 gezeigt. Es wird dabei von einer typischen Vergabegröße von /48-Netzwerken an Endkunden ausgegangen [30]. Die ersten drei Präfixbits werden von der Anzahl der Netzwerkbits (48) und der Anzahl der Subnetzbits (64) abgezogen. Wie die

	HD-Ratio 0.8	Ohne HD-Ratio
Netzwerke	70 B	35 T
Subnetze	490 T	2 QQ

Tabelle 2.2: Vergleich IPv6-Global-Unicast-Adressraum mit und ohne HD-Ratio (Einheiten: 1B=1E9, 1T=1E12, 1Q=1E15, 1QQ=1E18) [31]

Zahlen in Tabelle 2.2 zeigen, ergäben sich bei einer HD-Wert von 0.8 ein Schwellwert von 70 Milliarden Netzwerken und 490 Trillionen Subnetzen ab dem der IPv6-Adressraum als aufgebraucht erklärt werden würde. Das sind rund 897 Millionen /48-Netzwerke pro aktuell lebenden Menschen auf der Erde [32]. Die hierarchische Struktur beschränkt jedoch den Suchraum für aktive Netzwerke. Jede RIR verfügt über einen /12-Block von der IANA, aus diesem werden /29 bis /32 an ISPs vergeben, welche /48-Blöcke an einzelne Unternehmen verteilen [30, 33, 34]. Für das folgende Beispiel wird angenommen, dass ein ISP über eine /32-Allokation verfügt, aus denen er Adressen verteilen kann. Erst wenn für diese Allokation, der von der zuständigen RIR vorgegebene HD-Wert überschritten wird, kann ein weiterer Adressblock beantragt werden [34]. Mit einer /32-Allokation verfügt ein ISP über 65536 /48-Netzwerke zur Weitervergabe.

HD-Ratio	0.8	0.94
Schwellwert	7131	33689
Ausnutzung	0.109	0.514

Tabelle 2.3: ISP mit /32-Allokation und Vergabegröße von /48 – HD-Ratios [35]

Wie Tabelle 2.3 zeigt, ist in diesem Fall der Schwellwert mit einem HD-Ratio von 0.8 sehr niedrig. Von den 65536 möglichen /48-Netzwerken, gilt die Allokation als aufgebraucht, sobald 7131 Netzwerke vergeben wurden. Deswegen einigten sich die RIRs sehr schnell auf ein neues HD-Ratio von 0.94 [34, 33]. Der Schwellwert beträgt mit 33689 fast das Fünffache des Schwellwerts bei einem HD-Wert von 0.8. Dennoch liegt die Ausnutzung immer noch nur bei

51.4%. Sollte der Schwellwert überschritten sein, ist es sehr wahrscheinlich, dass der ISP über einen weiteren Allokationsblock verfügt. Es wird davon ausgegangen, dass, sobald der Schwellwert erreicht wurde, die Vergaben aus diesem Block eingeschränkt werden. Sollten im Rahmen einer IPv6-Messung für alle möglichen Netzwerke ein Paket zurückgeschickt werden, kann davon ausgegangen werden, dass ein Filter im Einsatz ist. Auch nach der Erhöhung des HD-Wertes auf 0.94 wird durch den Schwellwert die Anzahl an möglichen Netzwerken innerhalb eines ISP Netzwerkes deutlich reduziert. Unterscheiden sich der Antworttyp für aktive und nicht aktive Netzwerke, kann der Status des Zielnetzwerkes daraus abgeleitet werden.

2.4 Internet-weite Messungen

Je nach Ziel der Messung wird zwischen mehreren Scan-Typen unterschieden. Bei den Zielen kann es sich um einzelne Netzwerke, das gleiche Ziel in mehreren Netzwerken oder alle im Internet vorhandenen Adressen handeln. Während ein Großteil der Messungen auch in IPv6 möglich sind, kann bei Internet-weiten Messungen nicht mehr an jede Adresse ein Paket geschickt werden. Die Anzahl an Zielen muss reduziert werden und neue Tools wurden entwickelt. In diesem Kapitel werden die unterschiedlichen Scan-Typen erläutert und Tools für Internet-weite Messungen in IPv6 miteinander verglichen.

2.4.1 Scan-Typen

Internet-weite Scans werden verwendet, um Informationen am Ziel-Host abzufragen. Ein Host kann über einen oder mehrere Services verfügen, welche auf Ports zur Verfügung gestellt werden. Schon 2003 unterschied Lee et al. zwischen drei Arten von Scans [36]:

- a) Vertikale-Scans - Ziel ist ein Host, aber mehrere Ports. Dadurch kann festgestellt werden, welche Services an einem Ziel verfügbar sind. Meist werden solche Scans auf die *Well-Known-Ports* 1-1023 beschränkt. Diese können sowohl in IPv4 als auch in IPv6 durchgeführt werden. Tools wie Nmap wurden bereits für vertikale IPv6-Portscans angepasst [37].
- b) Horizontale-Scans - Es wird der gleiche Port bei mehreren Hosts anvisiert. Ein typisches Szenario für solche Scans ist zum Beispiel das Erkennen des Ausmaßes einer Schwachstelle.
- c) Block-Scans - Scans visieren mehrere Ports bei unterschiedlichen Hosts an.

Während Vertikale-Scans sowohl in IPv4 als auch in IPv6 durchgeführt werden können, sind Horizontale-Scans und Block-Scans in IPv6 sehr limitiert. Solange die Anzahl der Hosts auf einzelne Netzwerke beschränkt ist, können diese auch in IPv6 durchgeführt werden. Internet-weite Horizontale-Scans oder Block-Scans sind in IPv6 nicht möglich. Der IPv4-Adressraum erlangt durch diese Scan-Typen ein sehr hohes Maß an Transparenz, denn Internet-weite Scans

können von jeder Person mit Zugang zu Internet und Hardware durchgeführt werden. Internet-weite Horizontale-Scans und Block-Scans wurden angepasst. Die Anpassungen erfolgen im Wesentlichen in zwei Richtungen:

- 1) Durch das Erstellen von Hit-Lists. Adressen werden hierbei aus unterschiedlichen Quellen, wie DNS, CDN Logs oder aus der Bitcoin API gesammelt und in einer Adressliste zusammengefasst [1, 2, 5, 38, 39]. Gasser et al. zeigten schon 2018 auf, dass Hit-Lists nicht nur einmalig auf ihre Aktualität überprüft werden sollten, sondern in regelmäßigen Intervallen gescannt werden müssen, um als Datengrundlage für weitere Scans sinnvoll genutzt werden zu können. Hit-Lists bergen weiters auch die Gefahr, Vollständigkeit vorzutäuschen. Es ist schwierig zu beweisen, dass eine Liste an Adressen das gesamte IPv6-Internet repräsentiert und nicht nur gewisse Teile davon.
- 2) Durch *Target-Generation*-Algorithmen, welche anhand aktueller Vergabemuster neue Adressen generieren. Das Anwendungsszenario für diese ist das Finden weiterer aktiver Hostadressen in einem Netzwerk. Beide Methoden, Hit-Lists und Target-Generation-Algorithmen, werden auch gemeinsam genutzt [5]. Dabei werden Hit-Lists auf Muster untersucht und anhand dieser weitere mögliche Adressen abgeleitet.

2.4.2 Tools für Internet-weite Messungen

Wie bereits erwähnt, ermöglichen Anpassungen von Nmap IPv6 Vertikale-Scans und Block-Scans einzelner Teilbereiche des Internets. Am häufigsten findet Nmap jedoch Einsatz bei Portscans von einzelnen Adressen oder der Host-Erkennung in Netzwerken [37]. Es wurden darüber hinaus auch neue Tools entwickelt. Scan6 ermöglicht es, aktive Hosts in einem IPv6 Netzwerk zu entdecken [40]. Im Bereich Internet-weiter Messungen gibt es deutlich weniger Lösungen. Die modularste Lösung bietet ZMAP, welches mit einer Paketrate von 1 Million Paketen pro Sekunde alle Adressen des IPv4-Internets innerhalb einer Stunde scannen kann [15]. Es erfolgten auch schon erste Anpassungen von ZMAP für IPv6. In Tabelle 2.4 werden für diese Arbeit relevante Tools anhand ihrer Use-Cases miteinander verglichen.

Tool	Input	Use Case
Nmap[37]	Netzwerk Range(s)	Vertikale, Block Scans einzelner Netzwerke
Zmap TU Muenchen[1]	IPv6-Adressen	Internet-weite horizontale Scans
Zmap Christoph Kukovic[41]	IPv6-Adressen	Horizontale-Scans einzelner Netzwerke
Zmap Bachelorarbeit [6]	Netzwerk-Range(s)	Horizontale-Subnetz-Scans
Scamper[42]	IPv6-Adressen	Topologie-Erkennung mittels Paris-Traceroutes
Yarrp[43]	IPv6-Adressen	Highspeed-Tracerouting

Tabelle 2.4: Scan-Tools und ihre Einsatzgebiete

Wie Tabelle 2.4 zeigt, wurde ZMAP bereits mehrmals adaptiert. Die Version von Christoph Kukovic fokussiert sich auf das Finden weiterer Hosts in einem Netzwerk. Dabei können mittels Adressmasken sowohl unterschiedliche Host-Adressen in einem Subnetz als auch in mehreren Subnetzen gescannt werden [41]. Die Adaption in meiner Bachelorarbeit fokussiert sich auf das Scannen von Subnetzen [6]. Dabei wird in jedes Subnetz ein Paket geschickt. Die Interface-ID der Anfragepakete wurde dabei statisch auf ::1 festgelegt. Für Horizontale-Scans mehrerer Netzwerke eignet sich die Adaption von ZMAP der TU München. Damit können Listen an IPv6-Adressen gescannt werden [1]. In IPv6 können damit Hit-Lists gescannt werden [1]. Die Adaption der TU München erzeugt die Adressen nicht selbst. Zieladressen müssen vorab erstellt und anschließend dem Tool übergeben werden. Alle bisher beschriebenen Tools haben gemeinsam, dass pro Ziel-Host und Port ein Anfragepaket benötigt wird. Scamper und Yarrp sind Tools, welche mittels Tracerouting für die zweite Art an Internet-weiten Messungen eingesetzt werden, nämlich zur Topologie-Erkennung. Dabei wird der Fokus nicht auf den Ziel-Host gelegt, sondern auf die Route, die das Paket zum Ziel einschlägt. Tracerouting nutzt die Begebenheit, dass Hosts, welche den Time-to-Live-Wert eines Paketes auf 0 setzen, eine Time-Exceeded-Nachricht mit dem Code Hop-Limit-Exceeded an den Sender zurückschicken müssen[9]. Beginnend bei 1 werden die Time-To-Live-Werte der Anfragen erhöht und somit Time-Exceeded-Nachrichten der einzelnen Stationen, die das Paket passiert, erzwungen. Tracerouting braucht jedoch je nach Distanz zum Ziel eine Vielzahl an Paketen gegenüber herkömmlichen Scanmethoden.

3 Messaufbau

Das folgende Kapitel erklärt den Aufbau der Internet-weiten Messung. Zunächst wird der Ausgangspunkt der Messung beschrieben. Dabei wird erläutert, warum die Messung von einem und nicht mehreren Ausgangspunkten durchgeführt wird. Danach wird darauf eingegangen, was die Ziel-Hosts dieser Messung sind und wie sie über die RIRs verteilt sind. Abschließend wird die Messmethode genauer beschrieben.

3.1 Ausgangspunkt der Messung

Die Messung wird von einem Standort aus durchgeführt. Dabei wurde ein eigens für Messungen ausgelegter Server benutzt. Dieser befindet sich im Netzwerk eines österreichischen ISPs mit einem 1-Gigabit-Uplink. Der Server verfügte zum Zeitpunkt der Messung über Debian 9 als OS, einen Intel-Xeon E5-2520 als Prozessor und 32 Gigabyte an RAM. Die Spezifikationen für den Server wurden absichtlich gering gehalten, um zu zeigen, dass diese Messung unabhängig von verwendeter Hardware durchgeführt werden kann. Das einzige Kriterium für das Erreichen von höheren Scan-Raten ist das Verwenden einer mit dem Kernel-Modul PF_Ring-ZC kompatiblen Netzwerkkarte [44]. Mit dem Modul können auch die Kapazitäten eines 10-Gigabit-Uplinks ausgenützt werden. Bei Tracerouting werden oftmals mehrere Ausgangspunkte benutzt [42]. Bei der Topologie-Erkennung spielen die Ausgangspunkte eine wichtige Rolle, um zum Beispiel das Verhalten von Routen bei Startpunkten in unterschiedlichen Nationen zu beobachten. Die Route, die das Paket zu dem Ziel nimmt, wird im Rahmen dieser Messung nicht evaluiert, daher wird auf unterschiedliche Ausgangspunkte verzichtet. Der Time-to-Live-Wert der Anfragen wird auf das Maximum von 255 gesetzt, da garantiert werden soll, dass das Paket sein Ziel erreicht.

3.2 Zielnetzwerke

Basis der Messung sind geroutete IPv6-Netzwerke. Diese wurden am 19. Juli 2018 dem RIPE-RIS-Service entnommen [1]. Das Service besteht aus aktuell 24 auf der Welt verteilten Routing-Kollektoren, welche die Routing-Einträge als Rohdaten zu Verfügung stellen. In den Routing-Tabellen des RIS-Services befinden sich Präfixgrößen von /16 bis /128.

[1] https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris

Das Routing erfolgt anhand des *longest-prefix-match*, dabei wird der kleinste Routing-Eintrag für das Ziel gewählt. In Tabelle 3.1 werden die BGP-Einträge den IANA-Global-Unicast-Allokationen per *longest-prefix-match* zugeordnet. Die IANA-Allokationen teilen sich auf die einzelnen RIRs auf, sowie auf Blöcke für besondere Zwecke wie 6to4, einen Übergangsmechanismus für IPv6[22]. Die folgende Tabelle soll die geografische Verteilung der gerouteten Netzwerke zeigen.

Tabelle 3.1: Zuordnung der BGP-Präfixe zu IANA vergebenen IPv6-Blöcken

IANA-Allokation	Präfixe
6to4	1
AFRINIC	807
APNIC	15474
ARIN	17555
IANA	4
LACNIC	12769
RIPE-NCC	22801

In Tabelle 3.1 werden alle 69411 gerouteten IPv6-Präfixe gezeigt. Davon gehören die meisten Präfixe, mit einem Anteil von 32.85%, zur RIPE-NCC. Die zweitmeisten Einträge gehören mit 25.3% zu ARIN. Input für die Messung sind alle /32-Präfixe. In Tabelle 3.2 wird die Verteilung der 11487 /32-Präfixe gezeigt.

Tabelle 3.2: /32-Präfixe und zugehörige RIRs

RIR	Präfixe
AFRINIC	212
APNIC	1888
ARIN	1680
IANA	1
LACNIC	2875
RIPE-NCC	4781

Wie Tabelle 3.2 zeigt, ist von den /32-Präfixen der Anteil an RIPE-NCC Einträgen noch höher. Mit 4781 Präfixen gehören 41.62 % der /32-Präfixe zur RIPE-NCC. AFRINIC verfügt sowohl im globalen Datensatz als auch bei den /32-Präfixen über sehr wenige Einträge. LACNIC-Präfixe machen im Gegensatz zu den globalen BGP-Daten 25% aller /32-Präfixe aus. Alle /32-Netzwerke sind jedoch nicht der endgültige Input für die Messung. Für die Messung wurden größere Netzwerke wie z.B. ein /29 in acht /32-Präfixe aufgeteilt, da die Implementierung von ZMAP nur eine Liste an Netzwerken mit der gleichen Präfixgröße scannen kann. In der Input-Liste befinden sich 11487 /32-Präfixe.

Addiert man die größeren Präfixe, welche auf mehrere /32 aufgeteilt wurden, erhält man 33089 /32-Präfixe. In Tabelle 3.3 werden diese erneut den RIRs zugeordnet.

Tabelle 3.3: /32-Input-Präfixe für Scan und zugehörige RIRs

RIR	Präfixe
AFRINIC	368
APNIC	2204
ARIN	2557
IANA	1
LACNIC	3059
RIPE-NCC	24900

Tabelle 3.3 zeigt ein überraschendes Ergebnis. Während sich die Anzahl an Präfixen von AFRINIC, APNIC, ARIN und LACNIC nur sehr geringfügig geändert haben, ist die Zahl der RIPE-NCC-Präfixe erneut um ein Vielfaches gestiegen. Somit gehören ein Großteil der Präfixgrößen größer als /32 zur RIPE-NCC. Ob sich dies auf die Ergebnisse auswirkt, wird in Kapitel 5 analysiert. Die Entscheidung für /32-Präfixe basiert auf einer typischen Vergabegröße von /32-Blöcken an ISPs. Diese enthalten kleinere Netzwerke, die entweder zur Weiterallokation oder direkt an Unternehmen vergeben werden.

3.3 Messmethode

Die Methode für die Messung resultiert aus dem Ziel Internet-weit zu überprüfen, welche IMCPv6-Antworten benutzt werden. Dabei sollen alle oben beschriebenen Zielnetzwerke gescannt werden. Für dieses Szenario ist die in der Bachelorarbeit entwickelte Version von ZMAP die performanteste. Bei der ZMAP-Version der TU-München müsste ein neues Tool entwickelt werden, um eine Adressliste zu erstellen, die die Subnetze der /32-Präfixe abdeckt. Die Version in der Bachelorarbeit wurde explizit dafür entwickelt, Subnetze einer Liste an IPv6-Netzwerken zu scannen. Die Adaption wurde so programmiert, dass pro Subnetzwerk eine Adresse gescannt wird, bei der der Host-Identifier auf ::1 gesetzt wird. Der Sendealgorithmus wurde für diese Messung beibehalten. Im Fall, dass durch die Einschränkung auf ::1 die Zieldresse nicht aktiv ist, wird eine Error-Message oder keine Antwort erwartet. Im Fall, dass es sich um eine aktive Adresse handelt, sollte standardmäßig mit einem Echo-Reply geantwortet werden, außer es sind Filter im Einsatz, die Echo-Requests blockieren. Beide Fälle sind für diese Messung erwünscht. ZMAP generiert zufällige Adressen im Rahmen 2^{64-p}, wobei p die Präfixgröße darstellt. Somit stellt jede zufällige Adresse ein Subnetz dar. Pro Netzwerk müssten bei einem vollständigen Scan aller Subnetze 4,3 Milliarden Subnetze abgedeckt werden, was eine unvorstellbare große Anzahl an Zielen ist. Für die Evaluierung der Antworttypen ist es wichtig, das Antwortverhalten

von möglichst vielen Netzwerken zu betrachten und nicht von jedem einzelnen Subnetz dieser Netzwerke. Um jedoch zu sehen, ob sich das Antwortverhalten bei unterschiedlichen Subnetzen ändert, werden als Proof-of-Concept 86000 Subnetze pro Netzwerk gescannt. Dadurch ergeben sich 2,84 Milliarden Anfragepakete pro Scan-Durchgang. Es soll überprüft werden, ob das Scannen von 86000 anderen Subnetzen zu anderen Ergebnissen führt. Die gescannten Subnetze ergeben sich in der Adaption von ZMAP aus den Pseudozufallszahlen. Mittels eines Seeds können über mehrere Messungen die gleichen Subnetze gescannt werden. Standardmäßig wird pro Durchgang ein neuer zufälliger Seed gewählt. Durch das Angeben eines Seeds bei mehreren Messungen bleibt die Reihenfolge der Zufallszahlen und somit der gescannten Subnetze gleich. Um zu überprüfen, ob andere Subnetze pro Zielnetzwerk zu anderen Ergebnissen führen, werden zehn Messungen mit unterschiedlichen Seeds durchgeführt. Darüber hinaus soll festgestellt werden, welchen Einfluss verschiedene Scan-Geschwindigkeiten auf die Ergebnisse haben, da Router die Anzahl an Error-Messages innerhalb einer gewissen Zeitspanne limitieren. Pro Seed werden daher Messungen mit fünf verschiedenen Scan-Raten durchgeführt. Die ausgewählten Scan-Raten sind: 50000, 100000, 500000, 1 Million und 1.5 Millionen Pakete pro Sekunde. Bei minimaler Scan-Rate ergibt dies bei einer Echo-Request-Größe von 67 Bytes eine Auslastung von 3.35 Millionen Bytes pro Sekunde und bei maximaler Scan-Rate von 100.5 Millionen Bytes pro Sekunde. Damit ist der 1-Gigabit-Uplink bereits ausgelastet. Dies führt zu insgesamt 50 Scans, die im Rahmen dieser Messung durchgeführt werden. Die Messungen erfolgen in einem Zeitraum von zwei Monaten von September bis November 2018.

4 Anpassungen des Scan-Tools

Für diese Arbeit wurde als Scan-Tool die adaptierte Version von ZMAP aus meiner Bachelorarbeit gewählt. Während das Ziel in der Bachelorarbeit die Entwicklung eines Tools für Internet-weite IPv6-Messungen war, liegt der Fokus in dieser Arbeit auf der Evaluierung der verschiedenen ICMPv6-Antworttypen. In der Definition der Messmethode wurde die Anzahl an Scan-Durchgängen auf 50 festgelegt. Daraus ergibt sich Erweiterungsbedarf für die Adaption von ZMAP. Das Scan-Tool soll bei längeren Scan-Durchgängen unterbrochen werden können, um Serverwartungen durchzuführen. Zudem soll das Logging der Antworten optimiert werden um möglichst wenig Speicher in Anspruch zu nehmen. In diesem Kapitel wird beschrieben wie diese Anforderungen in ZMAP umgesetzt wurden. Zunächst wird der implementierte Sendealgorithmus aus der Bachelorarbeit analysiert. Dabei sollen Variablen festgestellt werden, die gespeichert werden müssen, um den Scan anschließend an der gleichen Stelle fortzusetzen. Im zweiten Schritt werden verschiedene Log-Formate miteinander verglichen und gezeigt, wie das Scan-Tool angepasst wurde, um die Log-Größe einer Antwort zu minimieren.

4.1 Interrupt-Funktionalität

ZMAP ist neben Masscan eines der wenigen Highspeed-Scan-Tools. Mit diesem können Millionen an Paketen pro Sekunde geschickt werden und somit die gesamte Bandbreite eines 1- oder 10-Gigabit-Uplink genutzt werden [15, 45]. Um eine zu hohe Auslastung der Zielnetzwerke zu vermeiden, sollte versucht werden die Anfragepakete gleichmäßig aufzuteilen. Folgend wird der Bachelorarbeit implementierte Sendealgorithmus analysiert. Dieser ist dafür verantwortlich, dass Schritt für Schritt ein Subnetz einer Liste an Netzwerken gescannt wird. Scans mit einer geringen Anzahl an Zielnetzwerken sollten dadurch nur mit niedrigen Scanraten durchgeführt werden. Der in der Bachelorarbeit implementierte Scanalgorithmus wird in Grafik 4.1 gezeigt. Dies ist notwendig, da ZMAP erneut adaptiert wird, um das Unterbrechen von längeren Scans zu ermöglichen.
Die in Grafik 4.1 gezeigte Schleife wird von jedem Thread so lange wiederholt, bis die Abbruchbedingung erfüllt ist. Diese ist erfüllt, sollte die maximale Anzahl an Zielen erreicht sein, alle Pakete empfangen worden sein oder keine neue Zufallszahl mehr abgefragt werden können. In einem Schleifendurchlauf startet jeder Thread bei dem ersten Netzwerk in der Input-Liste. Es wird eine Zufallszahl abgefragt und anschließend wird diese als Subnetz-ID in die

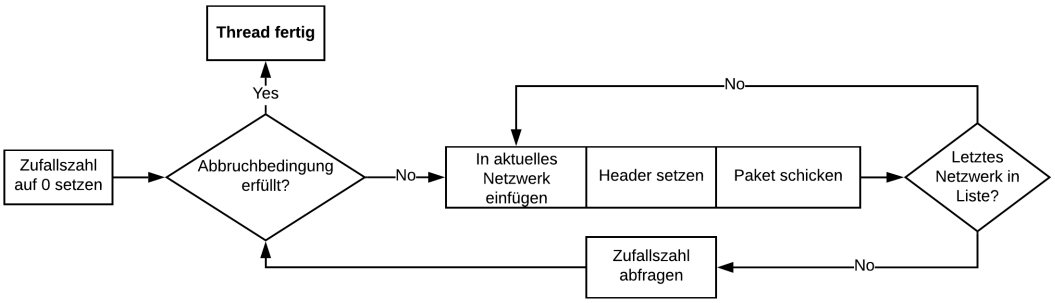


Abbildung 4.1: Sendealgorithmus für Subnetzscan in ZMAP [6]

Zieladresse eingefügt. Der *Host-Identifier* wird statisch auf die ::1 Adresse gesetzt. Ein Thread scannt dieses Subnetz bei jedem Netzwerk in der Liste und generiert erst danach eine neue Zufallszahl (=Subnetz). Die Implementierung für den Pseudozufallszahlengenerator ist für IPv6 und IPv4 gleich. Die Verwendung von zufälligen Subnetzen sowie zufälligen Adressen in IPv4 soll sicherstellen, dass sich diese bei jedem Scan ändern. Dies garantiert, dass auch bei höheren Scanraten nicht einzelne Netzwerke überlastet werden. Bei IPv6-Messungen kann es trotz der hohen Scanraten jedoch vorkommen, dass Scans mehrere Tage bis Wochen dauern. Bisher war es nicht möglich, ZMAP-Scans zu unterbrechen und beim aktuellen Stand fortzusetzen. Dies soll ermöglichen Serverupdates auszuführen oder Scankonfigurationen zu ändern, falls sich diese als nicht optimal erweisen. Nachfolgend wird gezeigt, wie ZMAP adaptiert wird, um den aktuellen Stand zu speichern und den Scan zu einem späteren Zeitpunkt wieder fortzusetzen. Die Schleife, in der Pakete geschickt werden, wird mit einem Interrupt-Handler versehen, welcher durch Strg+C ausgelöst wird, und somit als zusätzliche Abbruchbedingung in Grafik 4.1 zu sehen ist. Ein Vorteil dadurch ist, dass die Empfangseinheit von ZMAP wie bei anderen Abbruchbedingungen weitere zehn Sekunden auf Antworten wartet und nicht abrupt beendet. Die Sende-Threads sollen jedoch nicht nur unterbrochen werden. Damit ein Thread an der gleichen Stelle weitermachen kann, an der er geendet hat, müssen die Punkte in der folgenden Liste bekannt sein:

1. Nächste Zufallszahl (=nächstes Subnetz) sowie Zufallszahl, mit welcher der Sendethread aufhören kann. Für ersteres ist es wichtig, dass ZMAP unter Angabe eines *Seeds* ausgeführt wird. ZMAP wurde mit einem Pseudozufallszahlengenerator implementiert, einem deterministischen Algorithmus, der unter Angabe des Seeds beim nächsten Ausführen dieselben Pseudozufallszahlen liefert. Die Implementierung wurde mit unterschiedlichen Versionen von ZMAP schon öfters überarbeitet. In der Version, auf der die Adaption beruht, handelt es sich bei dem deterministischen Algorithmus um eine prim Restklassengruppe. Diese wird in Grafik 4.2 gezeigt. Grafik 4.2 zeigt, wie mehrere Threads die Restklassengruppe *durchschreiten*. Als Modul wird die nächstgrößere Primzahl gewählt als die maximale Anzahl an Subnetzen in der jeweiligen Präfixgröße der Input-Liste. Deswegen müssen alle Netzwerke in der Input-Liste über die gleiche Präfixgröße verfügen. Der Maximalwert des Moduls ist $2^{32} + 15$, welcher aus IPv4 übernommen wurde. In der Adaption können daher maximal /32-Netzwerke

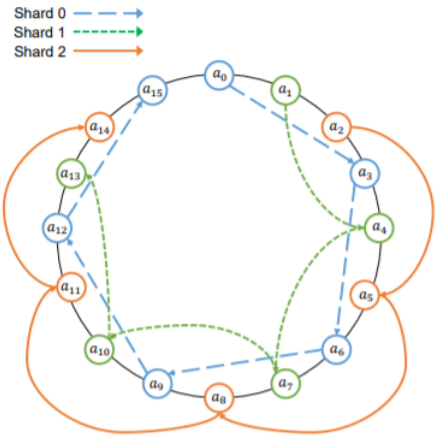


Abbildung 4.2: Sharding von ZMAP mit drei Threads (=Shards) [45]

gescannt werden, größere wie z.B. ein /29 müssen auf mehrere /32-Netzwerke aufgeteilt werden. Pro Scan wird eine neue Primitivwurzel gewählt und somit wird sichergestellt, dass die Reihenfolge der Subnetze pro Scan variiert. Durch Angabe eines Seeds erhält die Primitivwurzel, je nach Seed, immer den gleichen Wert. Im ersten Schritt muss daher der Seed (seed) des Zufallszahlengenerators gespeichert werden, um bei einer erneuten Initialisierung die prime Restklassengruppe die gleiche Reihenfolge an Pseudozufallszahlen zu garantieren. Wie Grafik 4.2 zu entnehmen ist, durchschreitet ein Thread die prime Restklassengruppe so lange, bis er den letzten Wert erreicht hat, der wiederum dem ursprünglichen Startwert entspricht (a0=a15,a14=a2,a13=a1). Daher müssen Anfangs- (first) und/oder Endwert (last) gespeichert werden, sowie die aktuelle Position in der primen Restklassengruppe (current).

2. Aktuelle Position in der Netzwerkliste, an der der Thread geendet hat.

   Basierend auf der Sendeschleife in Grafik 4.2, holt sich ein Thread pro Schleifendurchgang ein neues Subnetz zum Scannen und schickt an jedes Netzwerk in der Liste in das Subnetz ein Paket. Somit muss das letzte Zielnetzwerk (net_count) abgespeichert werden, was durch das Speichern der aktuellen Position in der Input-Liste passiert.

3. Anzahl der geschickten Pakete und Anzahl an maximalen Zielen insofern diese angegeben wurden. Die Anzahl der bereits geschickten Pakete (sent) könnte über die Position der Threads in der primen Restklassengruppe zurückgerechnet werden, wird aber der Einfachheit halber in der Interrupt-Datei als eigene Variable abgespeichert. Auch die Anzahl der maximalen Ziele (max_targets), welche als optionaler Parameter an ZMAP übergeben werden kann, muss in dem Fall, dass dieser Parameter gesetzt wurde, gespeichert werden.

Nach Auslösen des Interrupts wird von jedem Sende-Thread die notwendige Information, welche in obiger Aufzählung

mit Klammern versehen wurde, in die Interrupt-Datei geschrieben. Diese wird in Grafik 4.3 gezeigt. Bei erneutem Starten des Scanners wird diese eingelesen und alle notwendigen Variablen gesetzt.

```
seed 12345
id 0
#Shard Information
first 63461
last 63461
current 1300
sent 931
max_targets 458752
net_count 0
```

Abbildung 4.3: Variablen für Interrupt

4.2 Logging

Die Ausgabe von ZMAP erfolgt standardmäßig an Stdout. Sollte als Parameter -o übergeben werden, wird der Output in einer Datei gespeichert. Das Ausgabeformat ist standardmäßig eine csv-Datei. Zusätzlich können Felder angegeben werden, welche pro Antwort gespeichert werden sollen. Da Milliarden Anfragepakete ausgeschickt werden, führt dies auch bei einer sehr geringen Trefferquote zu einer großen Menge an benötigtem Speicher pro Antwortfeld, welches mitgeloggt werden soll. Die Anzahl der Antwortfelder sollte daher auf ein Minimum reduziert werden. Für diese Messung wurden die folgenden 3 Ausgabefelder gewählt. Die Ausgabefelder beinhalten die Herkunftsadresse,

```
--output-fields="saddr,classification,orig-dest-ip"
```

Abbildung 4.4: Output Varaiblen

die Klassifizierung des Antworttyps und die ursprüngliche Zieladresse des Anfragepaketes. Bei Echo-Replies sind Herkunftsadresse und ursprüngliche Zieladresse gleich, deswegen wurde die ursprüngliche Zieladresse durch eine Raute ersetzt. Bei Fehlernachrichten wird das ursprüngliche Anfragepaket an die Antwort angehängt, dadurch kann das ursprüngliche Ziel des Anfragepaketes extrahiert werden. Dies bedeutet, dass pro Antwort eine zweite 128 Bit lange IPv6-Adresse mitgeloggt wird. Die meisten Antwortadressen, die während der Messung empfangen wurden, haben jedoch nicht alle Bits gesetzt. Diese können im Textformat reduziert werden, beziehungsweise können 16 Bit Blöcke der Adresse, welche auf 0 gesetzt sind, durch einen doppelten Doppelpunkt abgekürzt werden. Im Rahmen von Testdurchläufen wurde festgestellt, dass die Annahme stimmt, denn es wurde eine durchschnittliche Log-Größe

einer IPv6-Adresse von 41.2 Bytes festgestellt. Der Unterschied zum Binärformat fällt dadurch deutlich geringer aus, denn binär würde eine IPv6-Adresse 32 Bytes an Speicher brauchen. Es wurde daher entschieden, das Logging für die gesamte Messung im Textformat zu belassen. Ein weiterer Vorteil vom Logging im Textformat ist, dass die Daten anschließend direkt ausgewertet werden können, ohne diese konvertieren zu müssen. Die Log-Größe der Klassifizierung wurde auf 1 Byte reduziert. Dabei wurden die Antworttypen mit einem Buchstaben als Abkürzung versehen. Diese werden in Tabelle 4.1 aufgelistet.

Tabelle 4.1: ICMPv6 Klassifizierung in ZMAP

Type	Code	Abkürzung für Klassifizierung
Destination Unreachable	No Route	r
	Admin Prohib	n
	Address Unreach	a
	No Port	p
	Ingress / Egress	y
	Reject Route	j
Time Exceeded	Hop Limit Exceeded	x
Echo Reply Message	Echo Reply	e
Anderer Typ		o

5 Evaluierung der Messergebnisse

Im Rahmen dieser Arbeit wurde eine Internet-weite IPv6-Messung aller gerouteten /32-Netzwerke durchgeführt. Bei der Messung sollen geroutete Netzwerke auf ihre ICMPv6-Antwortverhalten getestet werden. Es werden pro geroutetem /32-Netzwerk 86000 Subnetze gescannt. Es soll festgestellt werden, ob unterschiedliche Subnetze zu unterschiedlichen Ergebnissen führen. Je nach Seed werden andere Subnetze bei den Zielen gescannt. Deswegen wurden zehn Messungen mit unterschiedlichen Seeds durchgeführt. Dabei soll der Einfluss von unterschiedlichen Scan-Raten gemessen werden. Zu den getesteten Scan-Raten zählen 50K, 100K, 500K, 1M und 1.5M Pakete pro Sekunde. Die verwendeten Scan-Raten werden im Rahmen der Evaluierung als PPS angegeben. Pro Seed wurden die Scans mit allen fünf Scan-Raten wiederholt. Es soll evaluiert werden, welchen Einfluss Aliasing auf Echo-Replies hat. Dabei wird von dem Zielnetzwerk für jede Anfrage ein Echo-Reply zurückgeschickt, egal ob es die Zieladresse gibt oder nicht. Überdies soll gezeigt werden, woher die Antworten stammen, sollte es sich bei der antwortenden Adresse nicht um das Zielnetzwerk handeln. Netzwerke mit ähnlichen Antwortverhalten sollen erkannt werden und die Ursachen dafür gefunden werden. Je nach Ursache soll überprüft werden ob sich aus den erhaltenen Antworten der Status des Zielnetzwerkes ableiten lässt.

5.1 Evaluierungskategorien

Im ersten Schritt wird näher auf die verwendeten Kategorien während der Evaluierung eingegangen. Diese werden in Grafik 5.1 als Mengen visualisiert und anhand ihrer Größe miteinander verglichen.

BGP-Präfixe bezieht sich auf die am 18. Juli 2018 entnommenen Routing-Einträge. Sie stellen die Gesamtmenge an gerouteten IPv6-Netzwerken zu diesem Zeitpunkt dar. **/32-Zielnetzwerke** sind die Ziele dieser Messung und werden als Input an ZMAP übergeben. **Zieladressen** beziehen sich auf die 86000 Subnetze pro /32-Netzwerk, die von ZMAP je nach Seed zufällig ausgewählt werden. **Antworten** bezieht sich auf die Anzahl an Antworten, die während dieser Messung erhalten wurden. Im Rahmen von zwei Monaten wurde ein Terabyte an Antworten gesammelt. Die Menge an Antworten stammt von einer Untermenge **antwortender Adressen**. Vor allem bei Fehlernachrichten senden einzelne Gateways eine Vielzahl an Antworten zurück. Die Antworten lassen sich wieder in Netzwerke zusammenfassen, welche als **antwortende /32-Netzwerke** betitelt werden.

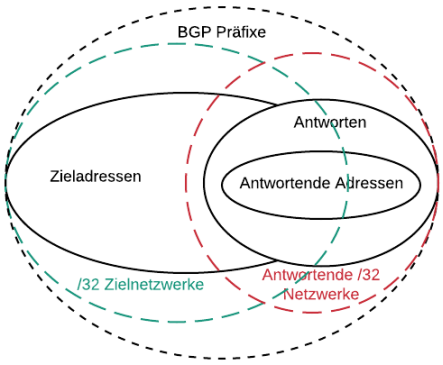


Abbildung 5.1: Ursprung Antworttypen

5.2 Vergleich Scan-Raten und Seeds

Die Ergebnisse der 50 Messungen werden für alle Seeds und Scan-Raten in Form von Tabelle 5.1 zusammengefasst. Für die zehn Seeds wird pro Scan-Rate das arithmetische Mittel gezeigt und die Standardabweichung in % vom Mittelwert. Die Ergebnisse werden anhand der in Grafik 5.1 visualisierten Kategorien ausgewertet.

Tabelle 5.1: Scanergebnisse zehn Seeds zu fünf Scan-Raten

Scan-Rate		50K	100K	500K	1M	1.5M
Antworten	$\overline{x}$	551.6M	421.7M	227.1M	168.5M	137.9M
	σ	1.22%	1.12%	1.83%	1.34%	1.18%
Antwortende	$\overline{x}$	17.0M	17.3M	16.9M	17.1M	16.7M
Adressen	σ	0.57%	1.58%	1.22%	1.22%	2.42%
Antwortende	$\overline{x}$	7087	7080.9	7067.1	7011.4	6705.8
/32-Netzwerke	σ	0.83%	0.75%	0.80%	1.20%	2.29%
/32-Ziel-	$\overline{x}$	19470.9	19335.5	18977.2	18555.9	18075.6
netzwerke	σ	1.24%	0.92%	0.71%	0.42%	0.35%

$\overline{x}$ ... Mittelwert von zehn Seeds

σ ... Standardabweichung in % vom Mittelwert

Seeds - Ein Seed gibt die Reihenfolge der Subnetze vor, welche gescannt werden. Durch die Beschränkung auf 86000 Subnetze, wurden je nach Seed unterschiedliche Subnetze gescannt. Innerhalb der Kategorien zeigen die Standardabweichungen für die verschiedenen Scan-Geschwindigkeiten nur sehr geringe Unterschiede und sind relativ klein. Sie reichen von einem Minimum von 0.35% bis zu einem Maximum von 2.42%. Unterschiedliche Seeds beeinflussen die

ausgewerteten Kategorien daher nur sehr gering. Die Scans wurden zu unterschiedlichen Tageszeiten durchgeführt. Schwankungen könnten ebenfalls auf zeitliche Unterschiede der Messungen zurückzuführen sein. Sollten andere Pakete zur gleichen Zeit wie die Messung das Zielnetzwerk erreichen, kann dies das Antwortverhalten der Netzwerke beeinflussen. Die Unterschiede bei antwortenden Adressen sowie /32-Netzwerke sind bei 1.5M Paketen pro Sekunde am größten, jedoch bei /32-Zielnetzwerken am kleinsten. Dies lässt auf Unterschiede durch die verkürzte Scan-Dauer bei Messungen mit hohen Scan-Raten schließen.

Scan-Raten - Es soll festgestellt werden, ob durch eine höhere Scan-Rate Informationen verloren gehen. Wie Tabelle 5.1 zeigt, ist die einzige betroffene Kategorie die Anzahl der Antworten. Während mit 50K Paketen pro Sekunde 551.6M Pakete empfangen wurden, wurden mit 1.5M nur noch 137.9M Pakete zurückgeschickt. Die Anzahl der antwortenden Adressen, antwortenden /32-Netzwerke sowie der /32-Zielnetzwerke reduziert sich nur geringfügig. Die Unterschiede zwischen den Scan-Raten wurden in Grafik 5.2 visualisiert.

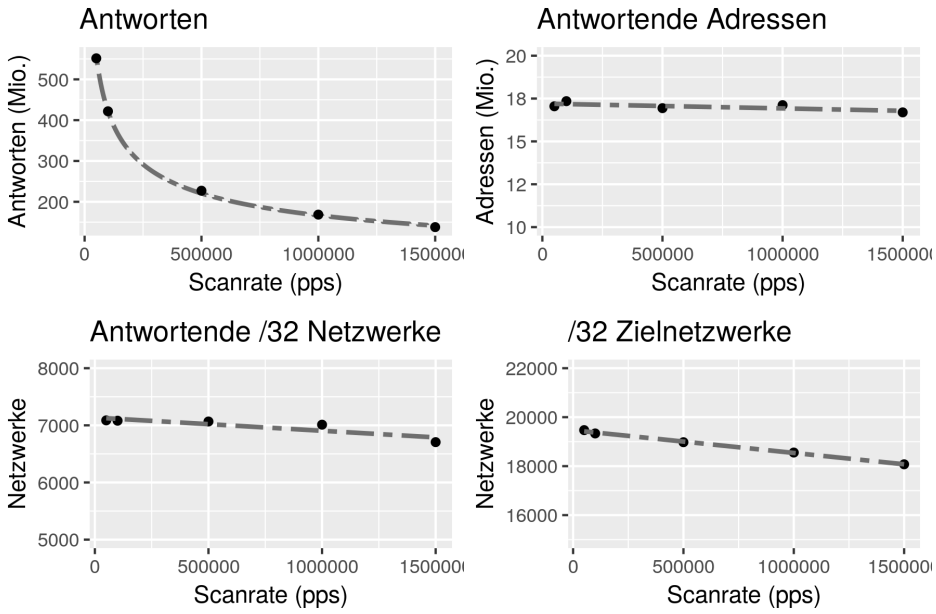


Abbildung 5.2: Einfluss von Scan-Raten auf Ergebnisse

Dabei entspricht in Grafik 5.2 jeder Punkt dem arithmetischen Mittel der zehn Seeds. Da die Standardabweichungen sehr niedrig sind, wurden sie aus Gründen der Leserlichkeit nicht visualisiert. Um einen funktionalen Zusammenhang der Daten zu visualisieren wird die lm - *linear-model*-Funktion in R benutzt. Der für jede Kategorie geschätzte funktionale Zusammenhang wird durch eine strichlierte Linie dargestellt. Grafik 5.2 zeigt, dass die Zahl der Antworten exponentiell abnimmt, die Zahl der antwortenden Adressen und /32-Netzwerke jedoch nur geringfügig linear. Die 551.6M Antworten bei 50K PPS entsprechen einer Trefferquote von 19%, das heißt ungefähr jedes fünfte Anfragepa-

ket führt zu einer Antwort. Mit einer Scan-Rate von 1.5M PPS beträgt die Trefferquote nur noch 4.5%. Es kommt zu einer Reduktion von 75% der Antworten gegenüber 50K PPS. Die Anzahl der antwortenden Adressen nimmt jedoch nur um 300K, also 1.8%, ab. Die antwortenden /32-Netzwerke nehmen um 5.4% und die /32-Zielnetzwerke um 7.2% ab. Je nach Wert der einzelnen Kategorien für einen Scan muss also entschieden werden, welche Scan-Rate die optimale ist. Scans, die antwortende Adressen fokussieren, sollten hohe Scan-Raten kein Problem bereiten. Der positive Effekt von 1M - 1.5M PPS ist die reduzierte Scan-Dauer. Für Scans, die auf möglichst viele Antworten setzen, sollte die längere Scan-Dauer in Kauf genommen werden und mit 50K oder weniger PPS gescannt werden.

Antworttypen - Die Scan-Ergebnisse werden weiter evaluiert, indem die Antworten in die einzelnen ICMPv6-Antworttypen gegliedert werden und der Einfluss von maximaler und minimaler Scan-Rate betrachtet wird. Zur Modellschätzung wird erneut die R-Funktion *lm* benutzt. Um die Antworttypen leichter unterscheiden zu können, wurde jedem Antworttyp in den Grafiken 5.3 und 5.4 ein Symbol zugeordnet. Diese werden als durch das Modell vorhergesagte Werte in den Grafiken angezeigt. In Grafik 5.3 wird der Einfluss von unterschiedlichen Scan-Raten auf die einzelnen Antworttypen und die Anzahl an Antworten visualisiert.

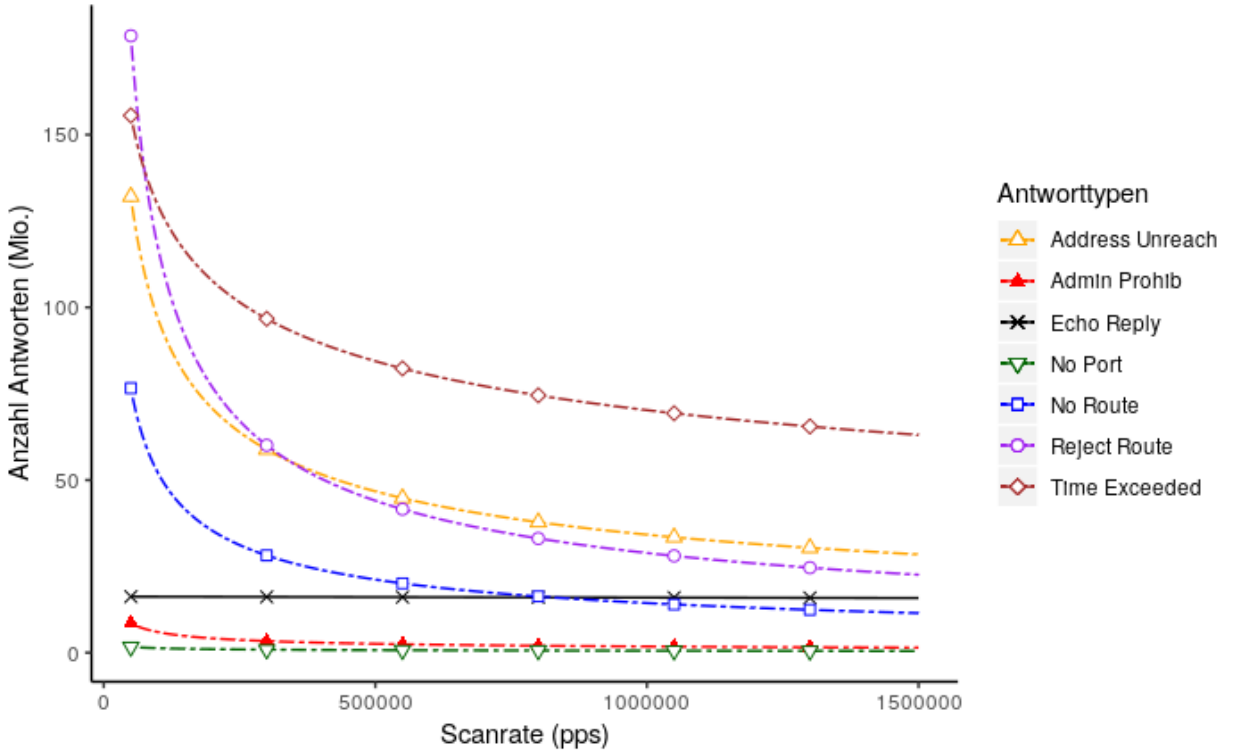


Abbildung 5.3: Antworten in Bezug auf Scan-Rate

Grafik 5.3 zeigt, dass sich Rate-Limiting unterschiedlich auf die einzelnen Antworttypen auswirkt. Reject-Route und

Address-Unreach Nachrichten sind dabei am stärksten betroffen. Echo-Replies bleiben konstant und die restlichen Fehlernachrichttypen werden durch höhere Scan-Raten weniger beeinflusst. In Grafik 5.4 wird die Auswirkung auf die Anzahl antwortender Adressen gezeigt.

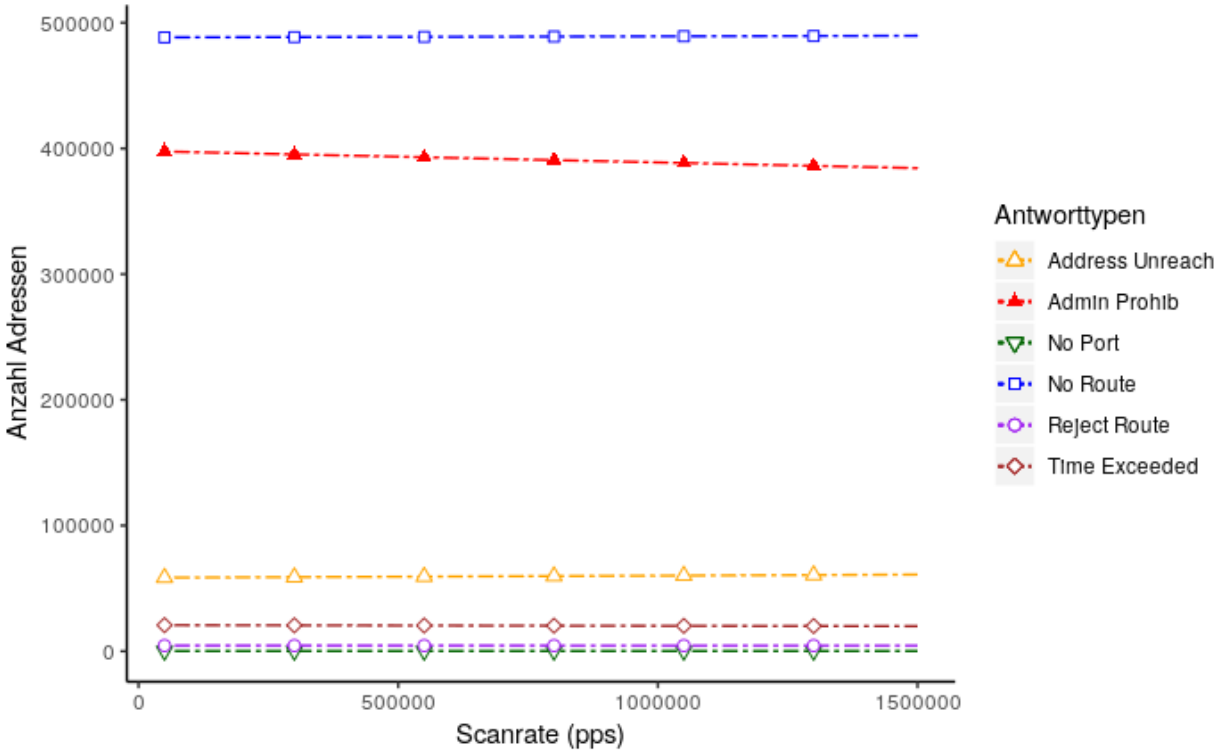


Abbildung 5.4: Adressen in Bezug auf Scan-Rate

In Grafik 5.4 wird ersichtlich, dass die Anzahl der antwortenden Adressen über alle Scan-Raten gleich bleibt. Echo-Replies werden zu Gunsten der Leserlichkeit nicht gezeigt, da sie mit 16.15M Antwortadressen die anderen Antworttypen bei weitem übertreffen. Echo-Replies-Adressen sind wie Echo-Replies-Antworten ebenfalls nicht von unterschiedlichen Scan-Raten betroffen. Kombiniert man die Grafiken 5.3 und 5.4, so wird ersichtlich, dass Antworttypen mit vielen Antworten und wenigen antwortenden Adressen, wie Reject-Route oder Address-Unreachable, stärker von Rate-Limiting betroffen sind. Mit 155M Paketen bei 1.5M PPS und 60.1M Paketen bei 50K PPS sind Time-Exceeded-Nachrichten der meist erhaltene Antworttyp. Time-Exceeded-Nachrichten stammen jedoch nur von einer verhältnismäßig sehr niedrigen Zahl an Adressen. Gleiches gilt für Reject-Route. Mit 165M Antworten von 3.6K Adressen ist das Verhältnis von Antworten pro Adresse noch größer. Im Gegensatz dazu stammen No-Route und Admin-Prohib-Nachrichten von sehr vielen verschiedenen Adressen. Bei Echo-Replies ist die Herkunfts- und Zieladresse immer gleich. Da bei Time-Exceeded und Reject-Routes immer wieder die gleichen Adressen antworten, dürfte ihr Ursprung näher am Ausgangspunkt des Scans liegen. Im Gegensatz dazu sollte der Ursprung von No-Route und Admin-Prohib

näher beim Ziel liegen. Während sich bei den meisten Antworttypen die Anzahl der antwortenden Adressen und Subnetze ähnelt, ist das bei Address-Unreachable-Nachrichten nicht so. Es dürfte Netzwerke geben, für die sehr viele Address-Unreach-Antworten zurückgeschickt werden und welche, für die das nur vereinzelt geschieht.

5.3 Aussortieren spezieller Netzwerkkonfigurationen

Netzwerke mit Konfigurationen wie Aliasing beeinflussen die Anzahl an Antworten stark, liefern aber keine Aussage über das Zielnetzwerk, außer, dass es sich um ein aktives Netzwerk handelt. Gleiches gilt zum Beispiel für Firewalling, wo für jede Anfrage mit dem gleichen Antworttypen geantwortet wird. Das Aussortieren soll anhand der Scan-Ergebnisse ermöglicht werden. Es wird für die weiteren Auswertungen ein Seed gewählt, da kaum Unterschiede zwischen unterschiedlichen Seeds festgestellt werden konnten. Da unterschiedliche Scan-Raten die Ergebnisse stark beeinflussen, wird das Aussortieren für jeweils 50K und 1.5M PPS gezeigt. Dies enspricht minimaler und maximaler Scan-Rate.

Ergebnisse vor dem Aussortieren - Bevor die Ergebnisse aussortiert werden, werden diese in Tabelle 5.2 gezeigt. Es wurden neue Kategorien für die Auswertung eingeführt. Das Zusammenfassen der Antworten basiert diesmal nicht auf /32-Ebene, sondern auf gerouteten BGP-Netzwerken. Die Netzwerke können dabei unterschiedliche Präfixgrößen haben. Es konnten jedoch nicht alle antwortenden Adressen in den BGP-Daten gefunden werden. Diese wurden in Tabelle 5.2 mit einer Klammer versehen. Es wird eine zusätzliche Kategorie Subnetze eingeführt. Diese beinhaltet /64-Netzwerke, für welche eine Antwort zurückgeschickt wurde.
Die in Tabelle 5.2 verwendeten Abkürzungen wurden in Kapitel 2 definiert. Wie sich in der Tabelle erkennen lässt, sind ein Großteil der Antworten Echo-Replies, verschiedene Destination-Unreachable-Codes und Time-Exceeded-Nachrichten. Was hervorsticht, ist die hohe Anzahl an Echo-Replies. Die Wahrscheinlichkeit, auf so viele aktive Adressen zu stoßen, ist jedoch sehr gering. Es wurden 86000 von 2^{32} möglichen Subnetzen pro Netzwerk gescannt. Es wurde eine (::1) von 2^{64} möglichen Host-IDs pro Subnetz gescannt. Dies lässt darauf schließen, dass einige Netzwerke Aliasing implementieren und ein weiteres Aussortieren der Antworten nötig ist, um Echo-Replies einen Wert zuweisen zu können.

Aliasing - Bei Aliasing antwortet ein Gateway für alle Hosts aus dem Netzwerk mit einem Echo-Reply, egal ob der Host aktiv ist oder nicht. Dies widerspricht RFC4443, in welchem definiert wird, dass ein Echo-Reply immer von der Zieladresse stammen sollte. Dieses Phänomen ist nicht neu, sondern wurde bereits bei mehreren Messungen als Problem erkannt [4, 2]. Es wurden bereits Lösungen vorgeschlagen, um Aliasing zu erkennen und auszusortieren. Alle Ansätze gehen davon aus, dass es in IPv6 unwahrscheinlich ist, auf aktive Adressen zu stoßen. Murdock et al. [4]

Tabelle 5.2: Scan-Ergebnisse je nach Antworttyp

Typ	Antworten	Antw. Adressen	Antw. Netze	Zielnetzwerke	Antw. Subnetze
Ergebnisse 1.5M PPS					
Echo-Reply	15.4M	15.4M	3.3K	3.3K	15.4M
No-Route	11.6M	484K (-139)	1.2K	2.7K	477.8K
Admin-Prohib	1.5M	382K (-20)	346	573	347.6K
Address-Unreach	26.9M	54K (-5K)	2.6K	5.5K	18.6K
No-Port	387.6K	181	29	29	179
Ingress / Egress	47	32	17	18	32
Reject-Route	20.1M	3.6K (-1K)	877	1.9K	2.9K
Time-Exceeded	60.1M	19.7K (-212)	2.4K	4.5K	16.3K
Ergebnisse 50K PPS					
Echo-Reply	16.15M	16.15M	3.8K	3.8K	16.15M
No-Route	76.5M	485K (-144)	1.1K	3.6K	475K
Admin-Prohib	7.9M	397K (-25)	346	592	362K
Address-Unreach	125.4M	53K (-5265)	2.6K	5.8K	18.5K
No-Port	13.7M	190	30	30	188
Ingress / Egress	46	31	17	19	31
Reject-Route	165M	3.6K (-962)	885	2K	2.9K
Time-Exceeded	155M	20K (-234)	2.4K	4.5K	16K

wählten als Ziel 3 zufällige Adressen in /96 Präfixen. Gasser et al. arbeiteten mit 16 kleineren Präfixen in gerouteten BGP-Präfixen und Präfixen der Hit-List mit mehr als 100 Adressen [2]. In dieser Arbeit wird eine neue Herangehensweise präsentiert, welche auf die bereits vorhandenen Scan-Ergebnisse angewandt werden kann. Ein Vorteil dieser Methode ist, dass keine zusätzliche Messung nötig ist und von variablen Präfixgrößen ausgegangen werden kann. Zur Erkennung von Aliasing werden folgende Parameter definiert. 1) k ist die Präfixlänge des Zielnetzwerkes. Ziel der Messung waren alle gerouteten /32-Netzwerke. Je nach Anzahl an Hops zum Ziel werden in erster Instanz mehrere kleinere Präfixe zu z.B. einem /32-Präfix aggregiert. Näher am Ziel werden diese in die tatsächlich gerouteten Präfixe aufgelöst. Es werden daher alle Echo-Replies per *longest prefix match* dem kleinsten gerouteten Netzwerk aus BGP zugeordnet. 2) m bestimmt die Anzahl an maximalen Zielen, welche in dem Fall der Anzahl der gescannten Subnetze entspricht. 3) p gibt die Präfixlänge der Liste an Netzwerken an, die ZMAP übergeben wurden. Im ersten Schritt wird P berechnet, was angibt, wie viele der ausgeschickten Pakete das Zielnetzwerk mit der Größe k erreicht hätten sollen. Dabei wird von der Gleichverteilung der Zufallszahlen von ZMAP ausgegangen. Stammen die Echo-Replies zum Beispiel von einem /33 Netzwerk, so wird bei einem p von 32 davon ausgegangen, dass $m/2$, also in dem Fall 43000

Pakete, dieses Netzwerk erreicht haben. Die Formel wird für größere Netzwerke geändert, da diese auf mehrere /32-Präfixe aufgeteilt wurden. Insgesamt erreichen somit $2^{p-k}m$ Pakete das Netzwerk. Diese Formel führt zum Beispiel bei einem *k* von 29 und einem *m* von 32 zu $8 * m$ Paketen, die das Zielnetzwerk erreichen sollten.

$$P(k,m) = \begin{cases} 0.5^{k-p}m & \text{if } k \geq p \\ 2^{p-k}m & \text{if } k < p \end{cases}$$

Ein Netzwerk wird als aliased gesehen, sollte die Anzahl an Echo-Replies *r* aus dem Netzwerk größer sein als der Schwellwert *s*. *s* ergibt sich durch die Multiplikation der Anzahl an an das Zielnetzwerk geschickten Paketen mit dem Gewicht *w*. Dieses repräsentiert die Schwierigkeit, eine aktive Adresse in IPv6 zu treffen.

$$s = \lceil P * w(k) \rceil$$

Für die Gewichtung wurde von einem HD-Ratio von 0.5 ausgegangen. Die ursprüngliche Grenze, ab der ein ISP eine neue Range beantragen musste, lag bei 0.8. Zur Wiederholung aus Kapitel 2, das HD-Ratio gibt an, wie weit ein Adressbereich bereits in Benutzung ist. Die Formel zur Berechnung des Schwellwertes für ein HD-Ratio ist HD-Limit $= 2^{(A-k)*\text{HD}}$ [34]. Das entspricht 7131 Subnetzen bei einer Allokationsgröße *A* von /48. Im Rahmen der Aliaserkennung wird für die Gewichtung von einem HD-Ratio von 0.5 und einer Allokationsgröße von /64 ausgegangen.

$$w = \text{HD-Limit}(0.5, k)/2^{64-k}$$

Berechnet man den Schwellwert für ein /32-Netzwerk, ergibt dies 86000 *P* Pakete, die in das Zielnetzwerk gelangen. Ein HD-Ratio von 0.5 ergibt bei einer Präfixgröße *k* von 32 ein HD-Limit von 65536 Subnetzen. Dies ergibt eine Gewichtung von $1.52587890625e-05$. Multipliziert man dies mit *m*, der Anzahl an ausgeschickten Paketen, erhält man einen Schwellwert *s* von 1.63, also aufgerundet 2. Mit 86000 gescannten Subnetzen werden /32-Netzwerke, die mehr als zwei Echo-Replies liefern, als aliased betrachtet. Bei Schwellwerten kleiner als Eins wird überprüft, ob nur eine Antwort von dem Ziel erhalten wurde. Ist dies der Fall, werden diese Antworten als *Non-Active* eingestuft. In Tabelle 5.3 wurde die Alias-Erkennung auf den ausgewählten Seed angewandt.

Tabelle 5.3: Ergebnisse Alias-Erkennung

		Aliased	Non-Aliased	Non-Active
1.5M PPS	Netzwerke	278	2666	392
	Adressen	15.39M	2779	392
50K PPS	Netzwerke	278	3082	423
	Adressen	16.15M	3215	423

Wie Tabelle 5.3 zeigt, sind bei 1.5M PPS 262 Netzwerke für 15.4 Millionen und bei 50K PPS 263 Netzwerke für 16 Millionen Echo-Replies verantwortlich. Es mussten rund 400 Netzwerke als Non-Active eingestuft werden, welche über einen Schwellwert kleiner 1 verfügten und nur eine Echo-Reply zurückschickten. Die Zahl der Non-Aliased Netzwerke ist mit 2682 bei 1.5 Millionen PPS und 3097 bei 50K PPS deutlich höher. Das Verhältnis von Netzwerken und antwortenden Adressen ist hierbei fast 1:1, was viel eher den Erwartungen entspricht. Addiert man die Zahl der Non-Active Netzwerken, wurden von allen gerouteten /32-Netzwerken somit nur von knapp 10% Echo-Replies erhalten. Aliased-Netzwerke wurden anhand ihrer Herkunft miteinander verglichen und den jeweils zuständigen RIRs zugeordnet. Grafik 5.5 zeigt die Herkunft von Netzwerken, welche Aliasing implementieren.

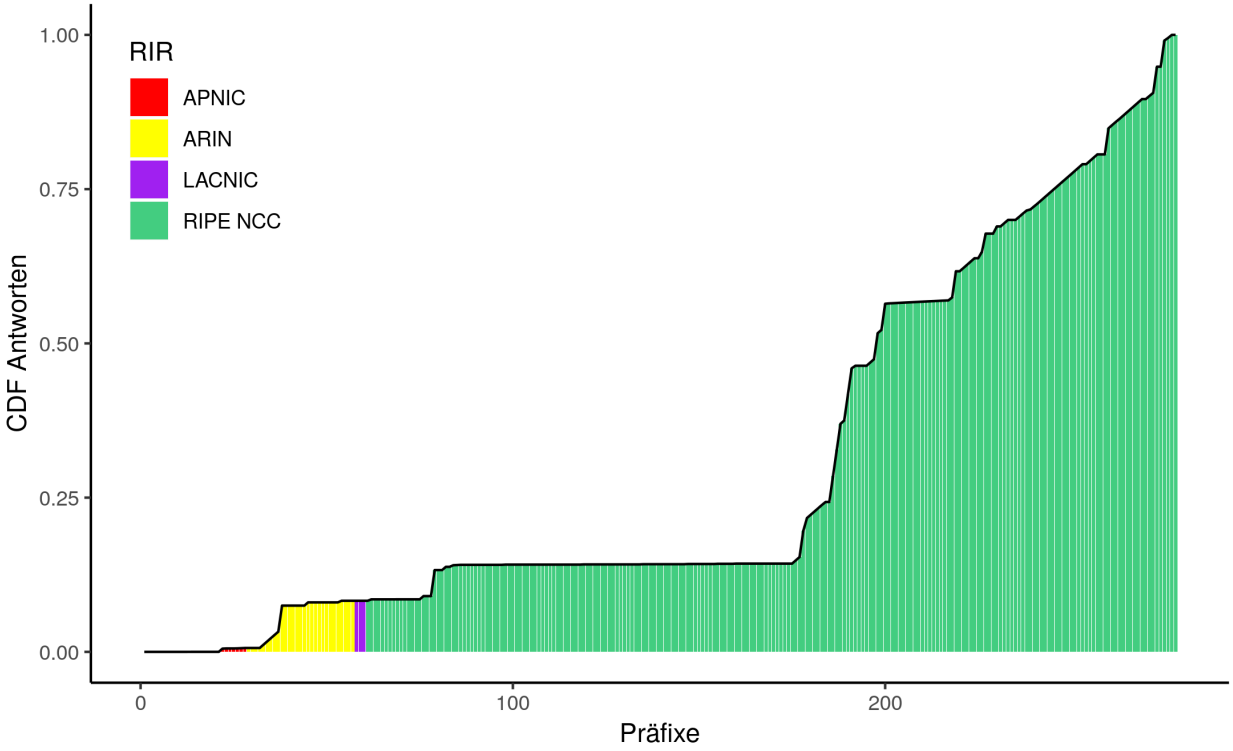


Abbildung 5.5: Präfixe mit Aliasing in Bezug auf RIRs

Was in Grafik 5.5 als erstes auffällt, ist, dass nur vier von fünf RIRs vertreten sind. Es wurde kein Netzwerk mit Aliasing in Adressbereichen von AFRINIC gefunden. Auch APNIC und LACNIC zeigen nur sehr wenige Netzwerke. Der erste Anstieg ist im ARIN-Adressbereich zu verzeichnen, auch wenn dieser sehr viel geringer ausfällt als bei RIPE-NCC-Adressbereichen. Ein Großteil der Aliasing Konfigurationen scheint Netzwerke aus Europa und Teilen Asiens zu betreffen.

Auffällige Netzwerke - Ein Großteil der Antworten sind ICMPv6-Fehlermeldungen. Dabei zeigen einzelne Netzwer-

ke eine unglaublich hohe Anzahl an Antworten und verschiedene Herkunftsadressen. Um die Antworttypen besser interpretieren zu können, sollen diese besonderen Konfigurationen aussortiert werden. Dabei erfolgt die Gliederung in nur zwei Kategorien, aktiv und nicht-aktiv. Für das Aussortieren wird eine 2-Sigma-Filterung angewandt.

$$S(a) = \begin{cases} Active & \text{if } a < \lceil \overline{a} + 2\sigma \rceil \\ Non - active & \text{if } a \geq \lceil \overline{a} + 2\sigma \rceil \end{cases}$$

Netzwerke, deren Anzahl an Adressen im Bereich 2-Sigma liegt, gelten als aktiv. Adressen von Netzwerken mit einer zu hohen Anzahl an verschiedenen Antwortadressen werden als nicht aktiv gewertet. In Tabelle 5.4 werden die Ergebnisse für jeden Antworttyp gezeigt. Zusätzlich wird das arithmetische Mittel der Anzahl an Herkunftsadressen und die Standardabweichung ausgewertet.

Tabelle 5.4: Ergebnisse 2-Sigma Regel

	Aktiv		Nicht-Aktiv		$\overline{a}$	σ
Type	Antw. Adr	Netze	Antw. Adr	Netze		
Ergebnisse 1.5M PPS						
No-Route	57K	1159	427K	1	417	12.6K
Admin-Prohib	30K	342	352K	4	1.1K	12.3K
Address-Unreach	22K	2601	32K	12	20.7	321.8
No-Port	51	28	130	1	6.2	24
Ingress / Egress	24	16	8	1	1.88	1.9
Reject-Route	2413	869	1193	8	4.1	22.1
Time-Exceeded	10.06K	2449	9606	3	8	168.1
Ergebnisse 50K PPS						
No-Route	56.8K	1144	428K	1	424	12.7K
Admin-Prohib	34.5K	342	362K	4	1.1K	12.8K
Address-Unreach	20.5K	2671	32.5K	13	19.8	315.8
No-Port	51	29	139	1	6.3	25.2
Ingress / Egress	24	16	7	1	1.8	1.6
Reject-Route	2367	876	1236	9	4.1	22
Time-Exceeded	10.1K	2483	9890	4	8	166

Die Ergebnisse in Tabelle 5.4 stimmen mit der Annahme, dass einzelne Netzwerke besonders viele verschiedene Adressen beitragen, überein. Die Anzahl an aussortierten Netzwerken ist mit 30 bei 1.5M PPS und 33 bei 50K PPS sehr gering und die Zahl der aussortierten Adressen sehr hoch. Es wurden 88% der No-Route und 92% der Admin-Prohib Adressen als nicht aktiv eingestuft. Time-Exceeded-Herkunftsadressen wurden zu jeweils rund 50% als aktiv und nicht

aktiv klassifiziert. Time-Exceeded-Antworten, welche als nicht aktiv kategorisiert wurden, stammen von 4 Netzwerken, aktive von 2483. Das Verhältnis von Netzwerken zu durchschnittlichen Herkunftsadressen bei Admin-Prohib ist mit 1.1K Adressen pro Netzwerk das höchste. Im Vergleich zu anderen Typen ist die Standardabweichung bei Netzwerken, die mit No-Route oder Admin-Prohib antworten, sehr hoch, während das arithmetische Mittel im Vergleich sehr niedrig ist. Dies deutet auf unterschiedliche Router-Konfigurationen in Bezug auf No-Route und Admin-Prohib hin.

Im nächsten Schritt wird der Beitrag der Antworttypen zu den Kategorien einzigartige Herkunftsadressen, Netzwerke und Subnetze betrachtet. Die Kategorie Subnetze entspricht wieder verschiedenen /64-Präfixen innerhalb des antwortenden Netzwerkes. Grundlage bilden Echo-Replies ohne Aliasing. Die Ergebnisse werden in Tabelle 5.5 sowohl in absoluten Zahlen als auch in Prozent dargestellt.

Tabelle 5.5: Beitrag der Antworttypen zu den Scan-Ergebnissen

Type	Antworten	Adressen	Antw. Netze	Antw. Subnetze
		Ergebnisse 1.5M PPS		
Echo-Reply	2779 - 0.0%	2779 - 2.24%	2666 - 39.99%	2778 - 2.73%
+ No-Route	10.4M - 9.16%	57030 - 45.95%	945 - 14.18%	50796 - 49.9%
+ Admin-Prohib	802K - 0.71%	30381 - 24.48%	174 - 2.61%	26536 - 26.07%
+ Address-Unreach	24.7M - 21.81%	21959 - 17.69%	1604 - 24.06%	14301 - 14.05%
+ No-Port	387K - 0.34%	47 - 0.04%	3 - 0.05%	41 - 0.04%
+ Ingress/Egress	35 - 0.0%	24 - 0.02%	2 - 0.03%	24 - 0.02%
+ Reject-Route	16.8M - 14.83%	2257 - 1.82%	309 - 4.64%	1515 - 1.49%
+ Time-Exceeded	60.1M - 53.15%	9640 - 7.77%	963 - 14.45%	5796 - 5.69%
Total	**113M**	**124K**	**6666**	**101.8M**
		Ergebnisse 50K PPS		
Echo-Reply	3215 - 0.0%	3215 - 2.54%	3082 - 44.37%	3082 - 2.94%
+ No-Route	69.9M - 14.42%	56811 - 44.85%	897 - 12.91%	50542 - 48.14%
+ Admin-Prohib	6.9M - 1.43%	34460 - 27.2%	171 - 2.46%	30712 - 29.25%
+ Address-Unreach	114.6M - 23.63%	20313 - 16.04%	1582 - 22.78%	13488 - 12.85%
+ No-Port	1.4M - 0.28%	47 - 0.04%	4 - 0.06%	41 - 0.04%
+ Ingress/Egress	35 - 0.0%	24 - 0.02%	1 - 0.01%	24 - 0.02%
+ Reject-Route	142M - 29.28%	2198 - 1.74%	287 - 4.13%	1473 - 1.4%
+ Time-Exceeded	150.1M - 30.95%	9610 - 7.59%	922 - 13.27%	5620 - 5.35%
Total	**484.9M**	**126K**	**6946**	**105M**

Die in Tabelle 5.5 gezeigte Auswertung ist rein quantitativ. Es erfolgt dabei keine Interpretation der einzelnen Antwort-

typen, sondern es werden nur die Zahlen in den jeweiligen Kategorien miteinander verglichen. Dies lässt rückschließen, welche Antworttypen von Router-Herstellern häufiger benutzt werden und welche nicht. Nachdem Netzwerke mit speziellen Konfigurationen aussortiert wurden, werden die Ergebnisse der einzelnen Antworttypen in der Reihenfolge, in der sie bisher beschrieben wurden, aufsummiert. Eine andere Reihenfolge würde die Ergebnisse verändern. Es wird überprüft, welchen Beitrag ein Antworttyp zu allen vorigen Antworttypen liefert. Während **Echo-Replies** in den Kategorien Adressen und Subnetze nur 2.24% beitragen, liefern sie knapp 40% der aktiven Netzwerke. Die guten Ergebnisse basieren auf aktuellen Vergabemustern, welche zu einer hohen Trefferquote von::1 Adressen in IPv6 führt. Bei einer Scan-Rate von 50K PPS erhöht sich die Zahl der Netzwerke auf 44.37%. Ändern sich aktuelle Vergabemuster nicht, bleiben Echo-Replies eine wertvolle Quelle für die Suche nach aktiven Netzwerken. Selbst bei einer schnellen Scan-Rate machen **Error-Messages** jedoch 99.99% der Antworten, 97.76% der Herkunftsadressen, 60.01% der Netzwerke und 97.27% der Subnetze aus. 60% der Netzwerke wären somit ohne Error-Messages nicht entdeckt worden, da diese nicht in den Echo-Reply-Netwerken enthalten sind. Der Umkehrschluss gilt nicht, denn Netzwerke, welche mit Echo-Replies geantwortet haben, können auch Error-Messages zurückgeschickt haben. **No-Route** liefert nach dem Aussortieren bei 1.5M PPS mit knapp 46% der totalen Adressen die höchste Anzahl an Herkunftsadressen und mit knapp 50% ebenfalls die meisten Subnetze. Unterschiedliche Scan-Raten wirken sich nur auf die Anzahl der Antworten aus. Die Anzahl reduziert sich von 70M bei langsamer Scan-Rate auf 10.4M bei 1.5M PPS. Anschließend werden die Antworten des Typs **Administratively-Prohibited** hinzugefügt. Diese stellen knapp ein Viertel aller einzigartigen Herkunftsadressen sowie Subnetze dar, welche noch nicht in No-Route und Echo-Replies enthalten sind. Unterschiedliche Scan-Raten wirken sich auf diesen Typ nur geringfügig aus. Zu diesen Ergebnissen kommt ein beachtlicher Anstieg mit den Antworten des Typs **Address-Unreachable**. Sie sind neben Echo-Replies die zweitbeste Quelle für aktive Netzwerke. 1604 Netzwerke werden allein durch Address-Unreachable-Antworten beigetragen. Dies entspricht fast einem Viertel aller gesamt 6666 Netzwerke. Address-Unreachable liefert knapp 22K an zusätzlichen Herkunftsadressen, welche keine der vorherigen Antworttypen zurückgeschickt haben, sowie 14301 verschiedene /64-Präfixe, welche als Subnetze bezeichnet werden. Bei den anderen Antworttypen ist der Unterschied von Herkunftsadressen zu Subnetzen nicht so groß wie bei Address-Unreachable. Bei niedrigerer Scan-Rate nimmt die Zahl an Antworten deutlich zu, die Verteilung bleibt aber gleich. Address-Unreachable-Antworten nehmen mit 23.63% bei 50K und knappe 20% bei 1.5M PPS etwas mehr als ein Fünftel aller Antworten ein. Die Steigerung zu den bisherigen Ergebnissen fällt bei Typ **No-Port** etwas geringer aus. Von 51 antwortenden Adressen, siehe Tabelle 5.4, sind 47 in den bisherigen Ergebnissen nicht enthalten. Von den 28 Netzwerken sind es drei. Es stellt sich heraus, dass **Ingress/Egress-Policy** ein Antworttyp ist, der nur sehr wenig in Verwendung ist. Obwohl von 16 Netzwerken, die Ingress/Egress-Policy Nachrichten zurückschicken, bei 1.5M PPS nur zwei neu sind und bei 50K PPS nur eine neues dabei ist, sind alle 24 Herkunftsadressen bei beiden Scan-Raten neu. Bei 1.5M PPS wurden von 2413 **Reject-Route**-Adressen 2257 noch nicht durch andere Antworttypen abgedeckt. Mit 16.8M Antworten liefert dieser Antworttyp bei 1.5M PPS damit rund

15% der Antworten und 309 neue Netzwerke. Auch nach dem Aussortieren sind bei 1.5M PPS fast 50% der Antworten Resultate von Routing-Loops. Bei 1.5M PSS verdoppeln **Time-Exceeded** Nachrichten die Anzahl der Antworten. Sie stammen aus 2449 Netzwerken, wobei 922 noch nicht in den Ergebnissen vorhanden sind. Der Einfluss von unterschiedlichen Scan-Raten ist bei Time-Exceeded geringer als bei anderen Antworttypen. Dies lässt darauf schließen, dass Routing-Loops näher am Ziel stattfinden, da dort die Auswirkungen von Rate-Limiting geringer sind. Die Anzahl an neuen Subnetzen ist mit 5.8K bei 1.5M PPs im Gegensatz zu rund 9600 neuen Herkunftsadressen jedoch relativ niedrig. Daher wird die Annahme getroffen, dass die Herkunft von Code Hop-Limit-Exceeded sich größtenteils noch auf Gateways der globalen IPv6-Infrastruktur beschränkt und nicht auf die Zielnetzwerke selbst.

5.4 Antworten aus dem Zielnetzwerk

Liegt der Ursprung der Nachricht außerhalb des Zielnetzwerkes, wird davon ausgegangen, dass diese weniger Information über das Ziel enthält. Um den Ursprung von Error-Messages zu visualisieren, werden die gleichen Bits von Herkunfts und ursprünglicher Zieladresse ausgewertet.

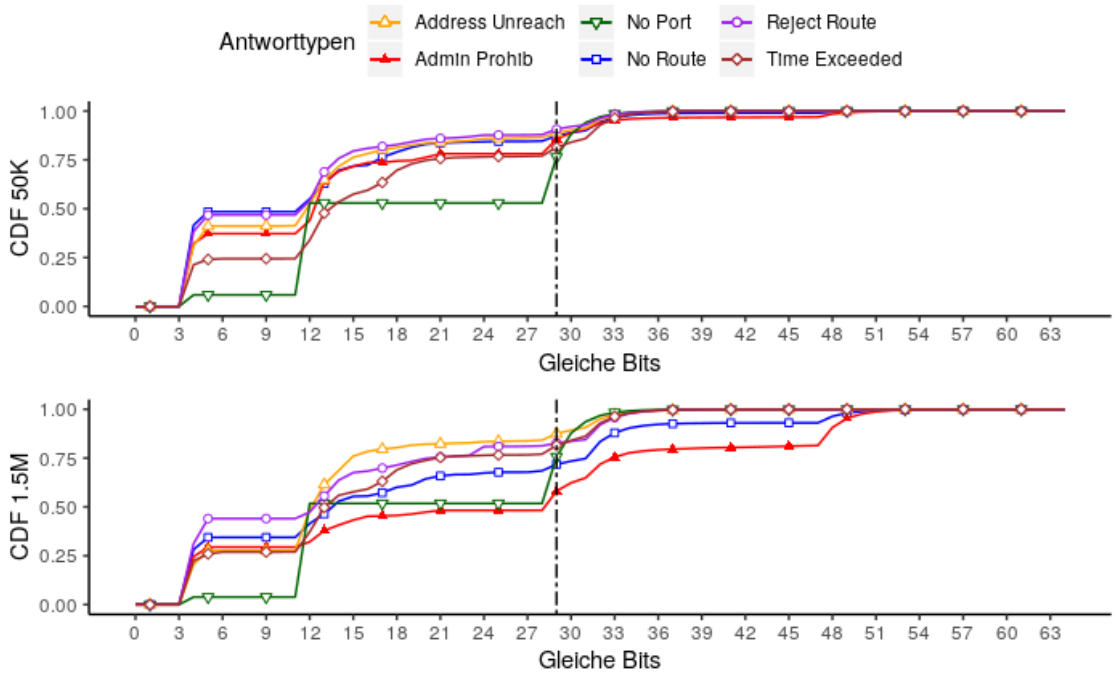


Abbildung 5.6: Bit-Vergleich Antworten

Grafik 5.6 enthält eine senkrechte Linie bei 29 Bits, die die Netzwerkgrenze symbolisiert, da Allokationen an ISPs typischerweise im Bereich von /29 bis /32 liegen. Für das Auswerten der gleichen Bits wurde die Herkunftsadresse mit

der ursprünglichen Zieladresse XOR gerechnet. Das erste Auftreten einer Eins im Ergebnis gibt das Ende der gleichen Bits an. Wie Grafik 5.6 zeigt, tritt der erste größere Anstieg an Antworten nach vier Bits auf. Dies ergibt sich aus den IANA-Allokationen an die einzelnen RIRs im Jahre 2006 [22]. Diese wurden in Tabelle 5.6 aufgelistet.

Tabelle 5.6: Verteilung von IPv6-Blöcken durch IANA 2006 (4 bis 8 gleiche Bits) [22]

IPV6 Block	RIR
2400:0000::/12	APNIC
2600:0000::/12	ARIN
2800:0000::/12	LACNIC
2a00:0000::/12	RIPE-NCC
2c00:0000::/12	AFRINIC

Wie Tabelle 5.6 zeigt, unterscheiden sich die Allokationen bereits in den ersten vier bis acht Bit. Kombiniert mit dem Anstieg in Grafik 5.6, der bei ebenfalls vier Bit verzeichnet wird, bedeutet dies, dass der Ursprung von bis zu 50% der Antworten eines Antworttyps in einem anderen Kontinent als das Ziel liegen kann. Davon sind bei einer langsamen Scan-Rate vor allem No-Route, Reject-Route, Address-Unreach und Admin-Prohib betroffen. Bei Time-Exceeded sind nur knapp 25% betroffen und No-Port zeigt fast keinen Anstieg bei vier Bits. Der zweite größere Anstieg beginnt bei zehn Bits. Die Vergabe der IANA basierte nicht nur auf Basis /12, sondern es wurden über die Jahre auch mehrere kleinere Allokationen gemacht. Tabelle 5.7 zeigt einige der kleineren Allokationen, welche acht bis zwölf oder zwölf bis 16 gleiche Bits aufweisen.

Tabelle 5.7: Verteilung von IPv6-Blöcken durch IANA (8 bis 12 / 12 bis 16 gleiche Bits) [22]

IPV6 Block	RIR
2610:0000::/23	ARIN
2620:0000::/23	ARIN
2630:0000::/12	ARIN
2a10:0000::/12	RIPE-NCC
2002:0000::/16	6to4
2003:0000::/18	RIPE-NCC

Da RIRs in den in Tabelle 5.7 gezeigten Kategorien öfters vertreten sind, ist bei 10 gleichen Bits oder mehr die geografische Trennung der Antworten nicht mehr so einfach möglich. Bis zur Netzwerkgrenze bei 29 Bits nimmt die Zahl der Antworten stetig zu. Bei 50K PPS liegen ungefähr 75% der antwortenden Adressen außerhalb des Zielnetzwerkes. Bei 1.5M PPS liegen sogar mehr Antworten innerhalb des Zielnetzwerkes, da Antworten von außerhalb stärker

von Rate-Limiting betroffen sind. Bei Admin-Prohib steigt dadurch der Anteil an Antworten aus dem Zielnetzwerk auf bis zu 50%. Um genauere Aussagen zu treffen, werden die Antworten per *longest-prefix-match* den BGP-Netzen zugeordnet. Tabelle 5.8 zeigt die Anzahl an Error-Messages aus dem Zielnetzwerk.

Tabelle 5.8: ICMPv6-Fehlernachrichten aus dem Zielnetzwerk

Type	Antworten	Adressen	Zielnetzwerke
	Ergebnisse 1.5M PPS		
No-Route	2.65M	55K	601
Admin-Prohib	287K	24K	245
Address-Unreach	3.76M	17K	1801
No-Port	186.5K	49	26
Ingress / Egress	27	20	13
Reject-Route	3.4M	1180	439
Time-Exceeded	12.9M	5640	1713
Total (unique)	10.3M	97K	2549
	Ergebnisse 50K PPS		
No-Route	10.1M	54.7K	658
Admin-Prohib	13.4M	28K	247
Address-Unreach	15.9M	15.8K	1860
No-Port	644K	48	26
Ingress / Egress	26	19	13
Reject-Route	19M	1180	437
Time-Exceeded	31.9M	5669	1730
Total (unique)	79M	105K	3549

In Tabelle 5.8 wird ersichtlich, dass bei 1.5M PPS nur 9% und bei 50K etwa 16% der Antworten aus dem Zielnetzwerk stammen. Was jedoch hervorsticht ist, dass die Anzahl an Antworten zwar geringer ausfällt, die Anzahl der Adressen im Verhältnis dazu aber sehr hoch ist. Als Beispiel wird No-Route genauer betrachtet. Tabelle 5.5 zeigt, dass bei 1.5M PPS 10.4M Antworten des Typs No-Route empfangen wurden, welche von 57030 Adressen stammen. Wie aus Tabelle 5.8 jedoch hervorgeht, stammen 55000 Adressen aus dem Zielnetzwerk. Diese sind für 2.65M Antworten verantwortlich. Die restlichen 7.75M No-Route-Antworten von außerhalb der Zielnetzwerke stammen von nur 2000 Adressen. Bei 50K PPS fällt der Unterschied noch stärker aus, da diese 2000 Adressen fast 60M No-Routes zurückschicken. Admin-Prohib ist weniger stark davon betroffen. Address Unreach, Reject-Route und Time-Exceeded zeigen ein ähn-

liches Verhältnis wie No-Route. Zusammenfassend lässt sich sagen, dass nur ein geringer Teil der Antworten aus dem Zielnetzwerk kommt.

5.5 Antwortverhalten der Netzwerke

Es soll herausgefunden werden, ob mehrere Netzwerke das gleiche Antwortverhalten zeigen. Dabei wird zwischen antwortenden Netzwerken und Zielnetzwerken unterschieden. Die Auswertung des Antwortverhaltens wird nur noch für 1.5M PPS durchgeführt, da bereits gezeigt wurde, dass sich unterschiedliche Scangeschwindigkeiten nur auf die Anzahl an Antworten, nicht aber auf die Anzahl der antwortenden Adressen, Netzwerke und Zielnetzwerke auswirkt. Dafür werden bereits aussortierte Antwortpakete den jeweiligen Herkunftsnetzen beziehungsweise ursprünglichen Zielnetzwerken zugeordnet. Pro Netzwerk wird mitgezählt, wie viele Antworten welchen Typs empfangen wurden. Abbildung 5.7 zeigt das Input-Format für RapidMiner. Die Spalte *Netzwerke* wird als Label definiert, die restlichen

```
Netzwerk,r,n,a,p,y,j,h,u,x,o,e
```

Abbildung 5.7: Datenformat für Antwortverhalten

werden als Integer interpretiert. Es werden nur Antworttypen ausgewertet, von welchen mindestens eine Antwort empfangen wurde. Im Rahmen der Auswertung der Antwortverhalten werden die in Kapitel 4 definierten Abkürzungen für die Antworttypen benutzt. Mittels RapidMiner werden aus den Netzwerken Cluster mit gleichem Antwortverhalten gebildet. Als Clusterverfahren wird k-means eingesetzt. Dieses Verfahren versucht die quadratischen Abweichungen vom Mittelwert des jeweiligen Clusters zu reduzieren [46]. Dazu werden zunächst ein zufälliges oder durch heuristische Algorithmen definiertes Set an Clusterzentren gewählt. In den folgenden Schritten werden Antworten dem nächsten Cluster zugeordnet und dessen Mittelwert berechnet. Der Mittelwert wird anschließend als neues Clusterzentrum verwendet. Dieser Vorgang wird so lange wiederholt bis das Zentrum sich nicht mehr verschiebt oder die Anzahl der *max optimization steps* erreicht wird. Sie wurden bei dem standardmäßig eingetragenen Wert von 1000 Wiederholungen belassen. Das Prozedere wird für *max runs* von zehn wiederholt, bei welchen ein unterschiedliches Set an Startpunkten für die Cluster gewählt wird. Startpunkte werden in RapidMiner anhand des 2006 von Arthur et al. vorgestellten k-means++ Heuristikalgorithmus ausgewählt [47]. Die Cluster mit den niedrigsten Abständen zu den Mittelwerten werden ausgegeben [48]. Bei den Antworttypen handelt es sich um numerische Attribute, deswegen wird als Verfahren für k-means die euklidische Distanz gewählt. Die Varianz der einzelnen Antworttypen in Bezug auf die Antworten pro Netzwerk ist jedoch wie in Tabelle 5.4 ersicthlich, sehr hoch. Es besteht die Gefahr, dass k-means keine wertvollen Cluster liefert.

5.5.1 Herkunftsnetzwerke

In einem ersten Schritt werden Cluster anhand der antwortenden Netzwerke gebildet und analysiert. Tabelle 5.9 zeigt die Ergebnisse des Clusterns bei einem k von sieben. Das Clustern wird mit steigendem k mit einem Startwert von 2 durchgeführt, die Ergebnisse des Clusterns bleiben jedoch über alle k-Werte sehr ähnlich.

Tabelle 5.9: Cluster k=7 Herkunftsnetzwerke

Cluster	Netzwerke
Cluster 0	6367
Cluster 1	1
Cluster 2	5
Cluster 3	29
Cluster 4	3
Cluster 5	246
Cluster 6	15
Total	6666

Wie Tabelle 5.9 zeigt, werden 95% der Netzwerke dem Cluster 0 zugeordnet. Auch bei anderen Werten für k bleibt die Zuordnung von einem Großteil der Antworten zu einem Cluster gleich. Um zu analysieren, warum so viele Netzwerke diesem Cluster zugeordnet wurden, wird dieser in Tabelle 5.10 genauer gezeigt. Dabei werden die geometrischen Zentren oder auch Mittelwerte der einzelnen Cluster analysiert.

Tabelle 5.10: Cluster 0 *Centroids* mit k=7

Antworttyp	Mittelwert
r	1131.9294801319302
n	201.95712266373488
a	1453.249253965761
p	31.684309722004084
j	1919.002198837757
x	1018.2541228207947
e	0.42170566986021674

Wie in Tabelle 5.10 zu sehen ist, sind bei Cluster 0 alle Antworttypen mit sehr niedrigen Mittelwerten vertreten. Es dominieren zwar die Antworttypen No-Route, Adress Unreachable, Reject-Route und Time-Exceeded, es werden aber

auch Netzwerke mit einem einzigen Echo-Reply als Antwort diesem Cluster zugewiesen. Die von k-means gebildeten Cluster haben wenig Aussagekraft über unterschiedliche Netzwerkkonfigurationen. Daher wird das Input-Format für RapidMiner und das Clustern überarbeitet. Die Anzahl an Antworten pro Antworttyp wird in ein binäres Attribut umgewandelt. Das heißt, es wird nur noch ausgewertet, ob ein Netzwerk einen gewissen Antworttyp zurückschickt oder nicht. Ist dies der Fall wird das Attribut auf 1 gesetzt. Ist dies nicht der Fall, erhält es den Wert 0. In Abbildung 5.8 wird als Beispiel ein Herkunftsnetzwerk gezeigt, aus welchem nur No-Routes empfangen wurden. Netzwerke mit

```
Netzwerk,1,0,0,0,0,0,0,0,0,0
```

Abbildung 5.8: Binäres Datenformat für Antwortverhalten

großem Abstand vom Mittelwert beeinflussen somit die Cluster weniger, da alle Antworttypen unabhängig von der Anzahl an Antworten die gleiche Wertigkeit erhalten. Das Clustern wird nun erneut für Herkunfts- und Zielnetzwerke durchgeführt. Es erfolgen mehrere Durchgänge mit steigendem k, mit einem Startwert von 2. Im Gegensatz zu den vorherigen Clustern sind die Netzwerke nun stärker auf die Cluster verteilt. Zum Vergleich werden zunächst die Ergebnisse für die Cluster mit einem k von 4 und danach mit einem k von 7 gezeigt.

Tabelle 5.11: Cluster k=4 Herkunftsnetzwerke

Cluster	Netzwerke
Cluster 0	2361
Cluster 1	2089
Cluster 2	1093
Cluster 3	1123
Total	6666

Im Gegensatz zu den Clustern mit den ursprünglichen Daten sind die Netzwerke nun viel gleichmäßiger über die Cluster verteilt. Um die Cluster näher analysieren zu können, werden die Mittelwerte der Cluster gegenübergestellt. Es wird versucht, jedem Cluster ein Merkmal (=Clustertyp) zuzuordnen, das das Antwortverhalten der Netzwerke in dem Cluster am besten beschreibt.

Tabelle 5.12 zeigt, dass die Cluster sehr unabhängig voneinander sind. Der dominante Antworttyp in **Cluster 0** sind Echo-Replies. Da bei Echo-Replies die Herkunftsadresse gleich der Zieladresse ist, handelt es sich dabei eigentlich sowohl um Herkunfts- als auch Zielnetzwerke. Die Wahrscheinlichkeit das Cluster 0 Netzwerke auch andere Antworttypen zurückschicken ist sehr gering. **Cluster 1** Herkunftsnetzwerke schicken zum Großteil Time-Exceeded-Nachrichten zurück, in einigen Fällen jedoch auch Address-Unreachable-Nachrichten. Ein weiteres Merkmal von Cluster 1 ist, dass aus den enthaltenen Netzwerken keine No-Routes stammen. Dies steht im Gegensatz zu **Cluster**

Tabelle 5.12: Cluster *Centroids* mit k=4

Antworttyp	Mittelwerte			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
r	0.028	0.0	1.0	0.0
n	0.010	0.078	0.122	0.020
a	0.090	0.330	0.536	1.0
p	0.0008	0.003	0.016	0.0017
j	0.028	0.223	0.216	0.094
x	0.091	0.826	0.467	0.0
e	1.0	0.0756	0.134	0.0
Clustertyp	Echo	Time/Address	No-Route/Mixed	Address

2, bei welchem immer No-Routes zurückgeschickt werden. Von diesen Netzwerken stammen oftmals auch Address-Unreachable und Time-Exceeded-Nachrichten, sowie zu einem kleineren Teil auch Reject-Routes. **Cluster 3** besteht zum größten Teil aus Netzwerken, die mit Address-Unreachable antworten.
Das Clustern erfolgt mit steigenden k-Werten und die Ergebnisse werden mit $k = 4$ verglichen. Tabelle 5.13 zeigt die Clusteraufteilung mit einem k von 7.

Tabelle 5.13: Cluster k=7 Herkunftsnetzwerke

Cluster	Netzwerke
Cluster 0	2446
Cluster 1	1115
Cluster 2	852
Cluster 3	391
Cluster 4	552
Cluster 5	1023
Cluster 6	287
Total	6666

Tabelle 5.13 listet die Netzwerke pro Cluster auf. Es finden sich auch bei sieben Clustern ähnlich große Cluster wie bei vier Clustern. Es haben sich auch etwas kleinere Cluster wie 3,4 und 6 gebildet. Im nächsten Schritt werden erneut die Mittelwerte miteinander verglichen und Clustertypen daraus abgeleitet. Es soll verglichen werden, ob Cluster aus k=4 sich in den Ergebnissen mit k=7 wiederfinden. Zusätzlich wird überprüft, ob das Antwortverhalten der Cluster mit den

Vorgaben in RFC4443 übereinstimmt. Außerdem werden Untersuchungen über die Ursachen der Cluster gemacht. Wird auf den RFC Bezug genommen, bezieht sich dies in der Analyse der Cluster immer auf RFC4443, in welchem ICMPv6 definiert ist.

Tabelle 5.14: Cluster *Centroids* mit k=7

Antworttyp	Mittelwerte						
	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
r	0.027	0.0	1.0	0.557	0.0	0.0	0.080
n	0.048	0.013	0.063	0.306	0.039	0.008	0.027
a	0.078	1.0	0.456	0.936	1.0	0.0	0.0
p	0.001	0.001	0.009	0.033	0.0	0.009	0.003
j	0.021	0.087	0.0	0.989	0.0	0.050	1.0
x	0.091	0.0	0.380	0.843	1.0	1.0	0.0
e	0.959	0.0	0.106	0.278	0.210	0.0	0.01
Clustertyp	Echo	Address	No-Route	Mixed	Time Address	Time	Reject
RFC kompatibel	✓	✗	✓	✗	✗	✗	✓
Ursache	Restriktiv	Firewall	Routen Aggregierung	Mixed	Routing-Loops		Reject Routes

Tabelle 5.14 zeigt die Mittelwerte, den Clustertyp, sowie RFC-Kompatibilität und Annahmen für die Ursachen, die zu dem jeweiligen Clustertyp führen. **Cluster 0** verhält sich ähnlich wie k=4 und hat Echo-Replies als dominanten Antworttyp. Es befinden sich sogar mehr Netzwerke bei einem k von 7 in Cluster 0 als bei einem k von 4. Die Netzwerke mit solchem Antwortverhalten wurden als restriktiv eingestuft. Es werden Echo-Replies zurückgeschickt, aber die Anzahl an Netzwerken, die zusätzlich Fehlernachrichten zurückschicken, ist sehr gering. In IPv6 ist es deutlich schwieriger, auf aktive Adressen zu treffen, als in IPv4. Die Annahme ist, dass deswegen ICMPv6-Echo-Requests deutlich weniger gefiltert werden und ein Drittel aller Herkunftsnetzwerke mit Echo-Replies antwortet. **Cluster 1** beinhaltet ein Sechstel aller Herkunftsnetzwerke, die ausschließlich Address-Unreach zurückschicken. Dieses Verhalten wurde auch in Cluster 1 im Clustering mit k=4 beobachtet. Die Anzahl der Netzwerke in diesem Cluster hat sich jedoch um 974 reduziert. Cluster 1 Netzwerke schicken keine No-Route zurück, die auf inaktive Subnetze schließen lassen, sondern nur Address-Unreach. Das Verwenden eines einzigen Antworttypen lässt auf Filtern der Anfragen schließen. Gemäß RFC sollte Addresss Unreachable jedoch nicht für Firewalling eingesetzt werden. Ursprünglich sollte dieser über inaktive Adressen in aktiven Netzwerken informieren. Es werden jedoch 86000 zufällige Subnetze

pro Netzwerk gescannt. Die Wahrscheinlichkeit, dass inaktive Subnetze darunter sind, ist sehr groß. Inaktive Subnetze sollten jedoch No-Route-Antworten hervorrufen, welche von Cluster 1 Netzwerken nicht zurückgeschickt werden. Durch die Verwendung von Address-Unreachable für Firewalling wurde als nicht kompatibel mit dem RFC gewertet. Aus Netzwerken in Cluster 1 stammen auch andere Typen, die auf Firewalling deuten lassen. Dazu gehören die Antworttypen Administratively-Prohibited oder Reject-Route. **Cluster 2** beinhaltet 852 Netzwerke, aus denen in jedem Fall No-Route Antworten kommen, sowie zu einem geringeren Teil auch Antworten des Typs Address-Unreachable. Stammen aus dem Netzwerk neben No-Rotues auch Address-Unreachable Nachrichten, so ist die Wahrscheinlichkeit sehr hoch, dass kein Filter im Einsatz ist. Der Cluster stimmt mit Cluster 2 aus dem Clustering mit k=4 überein, lediglich die Anzahl an Netzwerken hat sich um 241 verringert. Die Vielzahl an No-Route Nachrichten ist auf zwei Ursprünge zurückzuführen: 1) Es gibt deutlich mehr Subnetze oder kleinere Netzwerke, die nicht aktiv sind. Existiert das Subnetz nicht, sollte für dieses eine Antwort des Typs No-Route zurückgeschickt werden. Ist das Subnetz aktiv, aber die Zieladresse nicht, sollte Address-Unreachable benutzt werden. Netzwerke, die beide Antworttypen zurückschicken, werden daher als *offen* eingestuft. 2) Routen-Aggregierungen von BGP könnten der Auslöser für zahlreiche No-Routes sein. Dazu trägt mitunter das Scansetup bei. Die Inputnetzwerke für diese Messung bestehen aus einer Sammlung aller 24 Routingkollektoren, da die Routingeinträge auf jedem der Kollektoren unterschiedlich sind. Die Ursache dafür ist das Aggregieren von Routen, um Routing-Tabellen möglichst klein zu halten. Dabei werden nach dem Prinzip des *Supernettings* kleinere Einträge durch einen größeren ersetzt. Das Vorgehen dabei wird in folgendem Szenario beschrieben: Ein ISP verfügt über ein /29-Netzwerk, aus denen er zwei /32-Adressblöcke auswählt. Diese benutzt er, um mit einer Vergabegröße von /48 kleinere Netzwerke an Endkunden zu vergeben. Anstatt die vergebenen Netzwerke einzeln an andere BGP Gateways zu kommunizieren, wird nur der /29-Eintrag weitergegeben. Für den Scan werden alle acht möglichen /32-Netzwerke, die sich aus dem /29-Eintrag ergeben, gescannt. Es sind von diesen acht Netzwerken jedoch nur zwei aktiv. Zusätzlich wurde aus den zwei aktiven /32-Netzwerken nur ein geringer Teil an Endkunden vergeben. Im Rahmen dieser Messung werden daher einige Anfragen in nicht aktive Adressbereiche geschickt, die die Vielzahl an No-Route-Antworten erklären könnten. **Cluster 3** enthält 391 Herkunftsnetzwerke, die sehr viele verschiedene Antworttypen zurückschicken. Es kann daher keine generische Aussage über diese Netzwerke getroffen werden, da sich unterschiedliche Konfigurationen auffinden lassen. Einerseits können Netzwerke mit Firewalling in diesem Cluster vorhanden sein als auch auch offene Konfigurationen wie in Cluster 2. Die RFC Kompatibilität musste jedoch verneint werden, da aus diesen Netzwerken oftmals Time-Exceeded-Nachrichten stammen. Das Hop-Limit der Anfragepakete wurde für diese Messung auf den Maximalwert gesetzt und kann daher nicht ohne Routing-Loop den Wert 0 erreichen. Dieser Cluster ist neu gegenüber k=4. **Cluster 4** ist mit 552 Netzwerken ebenfalls einer der kleineren. Basierend auf dem Antwortverhalten besteht er aus einer speziellen Gruppe von Netzwerken mit Time-Exceeded-Nachrichten und Address-Unreachable als dominanten Antworttyp. Diese Netzwerke sind sowohl für0 Routing-Loops verantwortlich, als auch von Firewalling wie in Cluster 1 betroffen. Dieser sowie der folgende

Cluster 5 mit 1023 Netzwerken, welche ausschließlich Time-Exceeded Nachrichten zurückschicken, wurden daher als nicht kompatibel mit dem RFC eingestuft. Als Ursache dafür wird die Fehlkonfiguration von Time-Exceeded-Nachrichten gesehen, da gemäß RFC, Nachrichten nicht über den selben eingehenden Link weitergeleitet werden dürfen. In diesem Fall muss, wie in Kapitel 2 beschrieben, Address-Unreachable verwendet werden. **Cluster 6** mit 287 Netzwerken besteht aus Netzwerken, die hauptsächlich für Reject-Route Antworten verantwortlich sind. Cluster 1, 4, 5 und 6 bei einem k von 7 bilden somit Subcluster von Cluster 1 und 3 aus k=4.

5.5.2 Herkunft Antwort-Cluster

Die Cluster werden nun genauer analysiert, indem sie den verantwortlichen RIRs zugeordnet werden. In einem ersten Schritt werden alle Herkunftsnetzwerke den Global-Unicast-Allokationen der IANA zugeordnet. In Tabelle 5.15 wird die Aufteilung auf die Allokationen abgebildet. Was in Tabelle 5.15 als Erstes auffällt, ist, dass zwei Herkunfts-

Tabelle 5.15: Antwortende Netzwerke in Bezug auf IANA-Allokationen

IANA-Allokation	Präfixe
6to4	1
AFRINIC	111
APNIC	972
ARIN	1017
IANA	1
LACNIC	1282
RIPE-NCC	3280
Total	6664

netzwerke fehlen. Diese müssen jedoch geroutet sein, da sie an den Ausgangspunkt der Messung zurückgeschickt wurden. Die verantwortlichen Antwortpakete werden in Abbildung 5.9 gezeigt. Die Anfragen in 2a01:8e01:xxxx und

```
2504:3200:100:46::2 x 2a01:8e01:a131:143::1
2504:3200:100:45::2 x 2001:4479:a131:143::1
```

Abbildung 5.9: Antworten aus nicht allokierten Adressbereichen

2001:4479:xxx resultieren in einer Time-Exceeded Nachricht von einer nicht in IANA Adressbereichen allokierten Adresse. Das heißt in BGP werden auch nicht allokierte Adressbereiche geroutet. Die zugehörigen Einträge in BGP sind 2504:3200:100:46::/126 und 2504:3200:100:45::/126. Wird die restliche Verteilung der Herkunftsnetzwerke über

die RIRs betrachtet in Tabelle 5.15, so stimmt diese mit der in Kapitel 4 gezeigten Verteilung der Input-Präfixe für diese Messung überein. Um zu sehen, ob es geografische Unterschiede gibt, die sich auf die Cluster auswirken, werden die Netzwerke je nach Cluster den einzelnen RIRs zugeordnet. Dabei wird die relative Verteilung der RIRs pro Cluster ausgewertet. Der Grundwert ist die Anzahl an Netzwerken pro Cluster. Der Prozentanteil sind die Netzwerke pro RIR je Cluster. Die Verteilung wird in Abbildung 5.10 präsentiert.

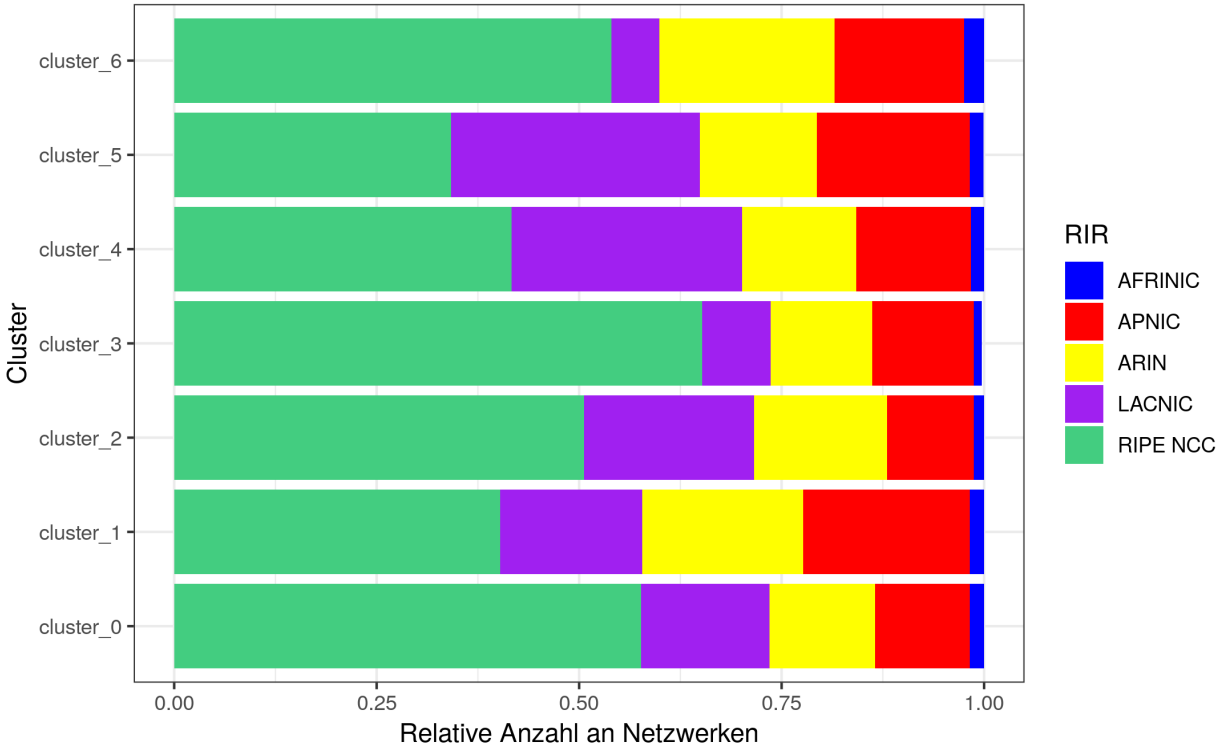


Abbildung 5.10: Herkunftsnetzwerke-Cluster in Bezug auf RIRs

Abbildung 5.10 beweist, dass sich die Ausmaße der in den einzelnen Clustern entdeckten Konfigurationen je nach RIR, also kontinental, unterscheiden. Ein Großteil der Ziele dieser Messung liegt in Adressbereichen der RIPE-NCC, dadurch ist es nicht verwunderlich, dass ein Großteil der Antworten ebenfalls aus RIPE-Netzwerken stammt. Für die Auswertung stehen jedoch die Zu- und Abnahmen der einzelnen RIRs im Gegensatz zu den anderen Clustern im Fokus.

- **Cluster 0 - Echo**, der Netzwerke beinhaltet, die Echo-Replies zurückschicken, hat mit 57% einen verhältnismäßig höheren Anteil an RIPE-Netzwerken als andere Cluster. Die anderen RIRs sind mit 11 bis 15% sehr gleichmäßig vertreten und heben sich nicht stark von anderen Clustern ab. Die Wahrscheinlichkeit, Echo-Replies zu erhalten, ist somit beim Scannen des ::1 Host-Identifiers in RIPE-NCC-Netzwerken deutlich höher als in ande-

ren.

- **Cluster 1 - Address-Unreach**, ist ein durch Firewalling hervorgerufener Cluster. Es ist gegenüber Cluster 0 eine deutliche verhältnismäßige Zunahme an ARIN- und APNIC-Netzwerken erkennbar. Address-Unreachable ist gemäß RFC4443 nicht für Firewalling vorgesehen. Anhand der Antwortverhalten dieser Netzwerke wird jedoch angenommen, dass der Antworttyp dafür eingesetzt wird. Der Anteil an RIPE-NCC-Netzwerken ist im Gegensatz zu Cluster 0 zurückgegangen. Dies korreliert mit der Annahme, dass einige RIPE-NCC offener bezüglich Echo-Requests reagieren. Der Anteil an LACNIC-Netzwerken hat sich kaum verändert.

- **Cluster 2 - No-Route**. Netzwerke dieses Clusters liefern vorrangig No-Routes als Antwort. RIPE-NCC trägt am meisten bei, verhältnismäßig mehr als in Cluster 1. In Kapitel 4 wurden die Zielnetzwerke ebenfalls den RIRs zugewiesen. Aus diesen und Cluster 2 leitet sich ab, dass, je mehr Zielnetzwerke aus einem RIR gescannt werden, desto mehr No-Routes von diesem stammen.

- **Cluster 3 - Mixed**. Dieser Cluster besteht aus Netzwerken, die oft mehrere Antworttypen zurückschicken. In diesem Cluster ist mit 65% der Anteil von RIPE-NCC zugehörigen Netzwerken besonders hoch. LACNIC weist im Verhältnis zu anderen Clustern weniger Netzwerke auf. ARIN und APNIC liegen mit 12% im Durchschnitt.

- **Cluster 4 - Time/Address** ist einer der zwei Cluster, der von Routing-Loops betroffen ist, wobei von diesem neben Time-Exceeded-Nachrichten auch Address-Unreachable stammen. Bei Cluster 4 ist mit 28,44% ein besonders hoher Anteil an LACNIC-Netzwerken erkennbar. Es ist einer der drei Cluster, bei denen RIPE-NCC-Netzwerke verhältnismäßig am wenigsten aufscheinen. Bei ARIN und APNIC mit 14% an Netzwerken ist kaum eine Änderung zu verzeichnen.

- **Cluster 5 - Time**. Wie Tabelle 5.14 schon zeigte, stammen aus diesem Cluster hauptsächlich Time-Exceeded Nachrichten. LACNIC verzeichnet wie in Cluster 4 mit 30.75 % einen überdurchschnittlich großen Anteil an Netzwerken. RIPE-NCC-Netzwerke machen 34.18% aus. In diesem Cluster ist auch bei Netzwerken aus APNIC-Adressbereichen ein leichter Anstieg von 14% auf 18% bemerkbar.

- **Cluster 6 - Reject** Netzwerke sind für Reject-Routes verantwortlich. ARIN ist in diesem Cluster mit 21.6% so stark vertreten wie in keinem anderen Cluster. LACNIC hingegen mit 5.9% so gering wie in keinem anderen. Der Anteil an RIPE-NCC Netzwerken ist mit 54% im Gegensatz zu Cluster 1, welcher auch durch Firewalling verursacht wird, etwas gestiegen. APNIC mit 16% und AFRINIC mit 2% sind im Vergleich zu anderen Clustern ebenfalls überdurchschnittlich vertreten. Der ARIN-Anteil ist sowohl in Cluster 1, welcher von Firewalling verursacht wurde, als auch in diesem Cluster besonders hoch. Dies lässt darauf schließen, dass Best-Practices bezüglich ICMPv6-Firewalling geographisch unterschiedlich sind.

5.5.3 Zielnetzwerke

Im nächsten Schritt wird das Clustern anhand der Zieladressen durchgeführt. Die Parameter für das Clustern, wie *max runs* und *max optimization steps* sind die gleichen wie bei den antwortenden Netzwerken. Nachfolgend werden in den Tabellen 5.16 und 5.14 die Ergebnisse für k-means mit einem k von 7 gezeigt. Die Cluster werden anschließend mit den sieben Herkunftsclustern verglichen.

Tabelle 5.16: Cluster k=7 Zielnetzwerke

Cluster	Netzwerke
Cluster 0	4396
Cluster 1	3278
Cluster 2	2187
Cluster 3	1336
Cluster 4	984
Cluster 5	339
Cluster 6	895
Total	13415

Die Aufteilung in sieben Cluster ist bei den Zielnetzwerken ebenso möglich wie bei den Herkunftsnetzwerken. Tabelle 5.16 zeigt die Anzahl der Zielnetzwerke pro Cluster. Der größte Cluster umfasst 32.77% aller Zielnetzwerke und der kleinste 2.53 %. Die Netzwerke sind wie die Herkunftsnetzwerke über alle Cluster verteilt. Zur Interpretation der Cluster werden in Tabelle 5.14 erneut die Mittelwerte gezeigt. Die Auswertung erfolgt anhand der gleichen Kategorien wie bei den Herkunftsnetzwerken, nach Clustertyp, RFC Kompatibilität und einer Annahme, welche Netzwerkkonfiguration zu diesem Cluster geführt hat.

Die Cluster in Tabelle 5.17 verfügen im Gegensatz zu den Herkunftsnetzwerken öfters über mehrere dominante Antworttypen. **Cluster 0** der Zielnetzwerke ist mit 4396 Netzwerken der größte. Für jedes dieser Netzwerke werden Address-Unreach-Antworten erhalten und bei einem Fünftel ebenfalls Time-Exceeded-Nachrichten. Andere Antworttypen sind kaum enthalten, Echo-Replies gar nicht. Dieses Verhalten entspricht typischerweise Firewalling. Wie bei den Herkunftsnetzwerken werden Anfragen fast zur Gänze mit einem Antworttyp abgehandelt. Address-Unreachable verliert deshalb in diesem Cluster seine Aussagekraft über das Ziel. Bei Firewalling kann nicht unterschieden werden, ob es sich bei dem Ziel um ein aktives oder nicht aktives Subnetz handelt. Address-Unreachable sollte gemäß RFC nur verwendet werden, wenn der Grund, warum das Paket zurückgeschickt wurde, mit keinem der anderen Antworttyp übereinstimmt oder um Routing-Loops zu vermeiden. Für Firewalling sollte Administratively-Prohibited verwendet werden. Dies wird in Cluster 0 nicht umgesetzt. Als weiterer Grund für die Nichtkompatibilität mit dem RFC zeigt ein

Tabelle 5.17: Cluster *Centroids* mit k=7

Antworttyp	Mittelwerte						
	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
r	0.098	0.061	1.0	0.001	0.0	0.047	0.130
n	0.032	0.003	0.004	0.002	0.015	0.991	0.062
a	1.0	0.0	0.0	0.0	0.0	0.0	1.0
p	0.003	0.006	0.0	0.0	0.0	0.008	0.007
j	0.061	0.025	0.006	1.0	0.0	0.070	0.101
x	0.194	1.0	0.0	0.0	0.0	0.079	0.280
e	0.0	0.123	0.090	0.137	1.0	0.0	1.0
Clustertyp	Address Time	Time Echo	No-Route	Reject Echo	Echo	Admin	Address Echo
RFC kompatibel	✗	✗	✓	✓	✓	✓	✓
Ursache	Firewall	Routing Loops	Routen Aggregierung	Firewall Restriktiv	Restriktiv	Firewall	Offen

Fünftel der Zielnetzwerke in Cluster 0 Routing-Loops. **Cluster 1** ist mit 3278 Netzwerken der zweitgrößte Cluster. Jedes der Netzwerke in Cluster 1 zeigt Routing-Loops. Als Ursache für diesen Cluster wird das Nichtbeachten des RFCs gesehen. Ein Achtel der Netzwerke in Cluster 1 antwortet neben Time-Exceeded Nachrichten auch mit Echo-Replies. Dies lässt auf eine gewisse Aktivität der Netzwerke schließen. Um genauere Aussagen treffen zu können, warum bei diesen Netzwerken Routing-Loops auftreten, wird mehr Information über Hardware und Netzwerkkonfigurationen im Einsatz benötigt. Dieser Herausforderung kann in zukünftigen Arbeiten begegnet werden. **Cluster 2** besteht aus 2187 Netzwerken und weist einen einzigen dominanten Antworttyp auf, nämlich No-Route. Für diese Netzwerke werden nur sehr selten andere Antworttypen erhalten. Als Grund für dieses Verhalten werden nicht-allokierte kleinere IPv6-Blöcke gesehen, für die keine Routing-Einträge vorhanden sind. Eine zweite Ursache könnte Firewalling der Netzwerke sein. Beide Erklärungen laufen jedoch darauf hinaus, dass es sich um nicht aktive Netzwerke handelt, beziehungsweise im Falle von Firewalling nicht zwischen aktiv und nicht aktiv unterschieden werden kann. Die Zielnetzwerke in **Cluster 3** haben als eindeutiges Merkmal den Antworttyp Reject-Route. Die Verhältnisse zwischen den Antworttypen in Cluster 3 ähneln denen in Cluster 1. Es wurden 1336 Netzwerke Cluster 3 zugeordnet. Eine mögliche Ursache für Reject-Route-Antworten kann Firewalling sein. Es könnten auch für einzelne Subnetze Reject-Routes definiert worden sein. Netzwerke, die nur Reject-Routes zurückliefern, deuten auf Firewalling hin. Netzwerke, die auch Echo-Replies zurückschicken, deuten auf den zweiten Fall hin. Dies stimmt mit der Beobachtung, dass ein

Achtel der Netzwerke in Cluster 3 mit Echo-Replies antworten, überein. Bevor Cluster 3 genaueren Scans unterzogen wird, sollten die Netzwerke also weiter aufgeteilt werden – in eine Gruppe, die Echo-Replies empfangen hat und in eine, in der das nicht geschehen ist. **Cluster 4** ist den Herkunftsnetzwerken von Cluster 0 sehr ähnlich. Für diese Netzwerke wurden meist nur Echo-Replies empfangen. In Ausnahmefällen wurden auch Administratively-Prohibited Antworten erhalten. Diese 984 Netzwerke sind Error-Messages betreffend sehr restriktiv. In diesen Netzwerken können aktive Subnetze nur durch Echo-Replies erkannt werden. **Cluster 5** ist ein kleinerer Cluster, für dessen Netzwerke hauptsächlich Administratively-Prohibited-Antworten erhalten wurden. Nur 339 Netzwerke fallen in dieses Verhaltensmuster. Administratively-Prohibited ist somit einer der weniger benutzten Antworttypen. Address-Unreachable ist in Verbindung mit Firewalling häufiger im Einsatz als der eigentlich dafür vorgesehene Antworttyp. Für eine genaue Untersuchung müsste das Verhalten bei aktiven ACLs von einzelnen Router-Hersteller beobachtet werden. Diese Aufgabe wird ebenfalls zukünftigen Arbeiten überlassen. **Cluster 6** beinhaltet 895 Netzwerke. Dieser Cluster wird als der wertvollste für Scans betrachtet. Neben Echo-Replies und Address-Unreachable werden auch andere Antworttypen zurückgeliefert. Diese Zielnetzwerke sind daher im Gegensatz zu den anderen Clustern sehr offen konfiguriert, was Error-Messages betrifft. Jedes der Netzwerke in Cluster 6 antwortet sowohl mit dem Antworttyp Echo-Reply als auch mit Address-Unreachable, was das Filtern von Anfragen ausschließen lässt. Address-Unreachable-Nachrichten können, sofern sie nicht für Firewalling verwendet werden, eine wertvolle Quelle für aktive Subnetze oder kleinere Netzwerke sein. Bei diesen muss im Gegensatz zu Echo-Replies keine aktive Adresse getroffen werden, um herauszufinden, dass es sich bei dem Ziel um ein aktives Netzwerk handelt.

5.5.4 Herkunft Ziel-Cluster

Die Zielnetzwerke, für die Antworten erhalten wurden, werden wie die Herkunftsnetzwerke den IANA-Global-Unicast-Allokationen zugeordnet. Im Gegensatz zu den Herkunftsnetzwerken gibt es für jedes Zielnetzwerk eine IANA-Allokation. In Tabelle 5.18 werden diese aufgelistet.

Tabelle 5.18: Herkunft Zielnetzwerke

IANA-Allokation	Präfixe
AFRINIC	165
APNIC	1879
ARIN	2185
IANA	1
LACNIC	3345
RIPE-NCC	5840
Total	13415

Es wird die in Tabelle 5.18 gezeigte Verteilung mit den Input-Netzwerken für den Scan verglichen. Ziel waren 24900 /32-Präfixe der RIPE-NCC, welche in 5840 aktiven Zielnetzwerken variabler Präfixlänge resultieren. Mit 3345 gehören die zweitmeisten aktiven Zielnetzwerke zur LACNIC. In der Input-Liste befanden sich 3059 /32-Präfixe der LACNIC. Der eine zur IANA gehörige Präfix resultierte auch in einem aktiven Zielnetzwerk. Bei der ARIN befanden sich 2557 /32-Präfixe in der Input-Liste, welche zu 2185 aktiven Netzwerken führten. Die 2204 /32-Präfixe der APNIC resultierten in 1879 aktiven Zielen, die 368 /32-Präfixe der AFRINIC in 165. In Abbildung 5.11 wird die Aufteilung der RIRs in den Ziel-Clustern visualisiert. Ein x-Wert von 1 entspricht allen Netzwerken in dem jeweiligen Cluster. Die RIRs werden in den gleichen Farben dargestellt wie bei den Herkunftsnetzwerken in Grafik 5.10 und der Aliasing Visualisierung in Grafik 5.5.

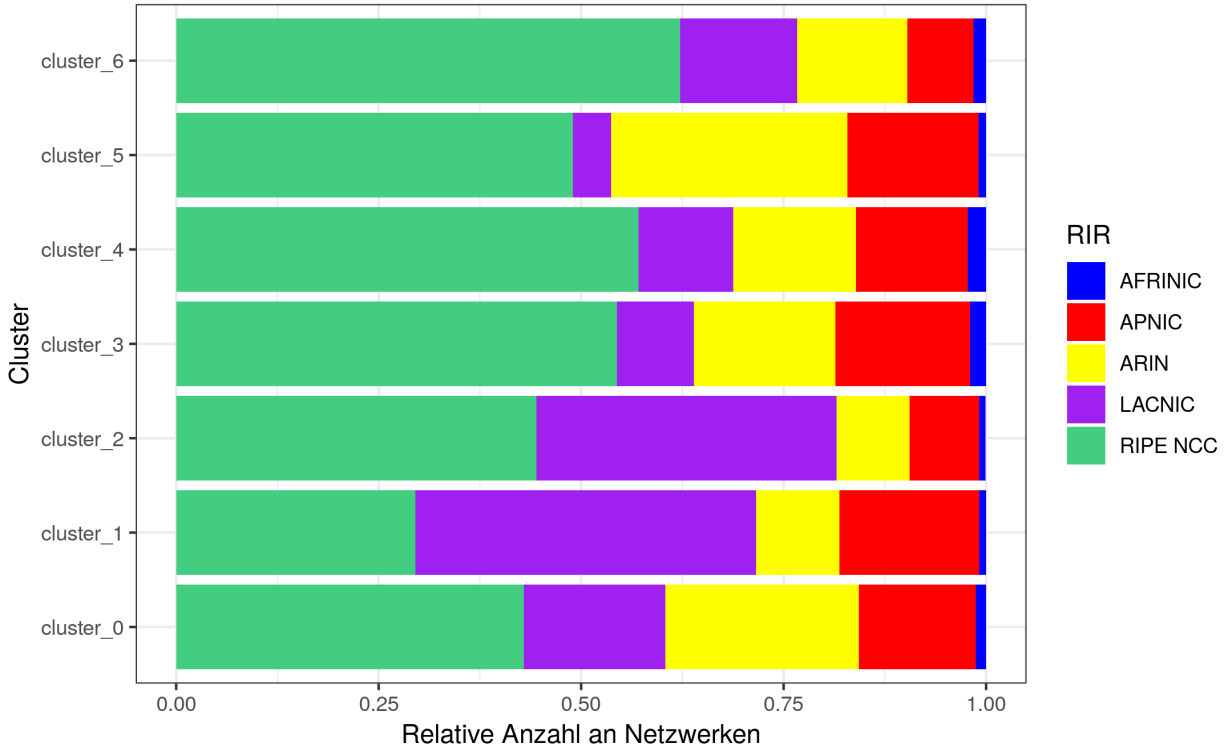


Abbildung 5.11: Zielnetzwerke Cluster in Bezug auf RIRs

In Abbildung 5.11 sind stärkere Unterschiede der RIRs über die Cluster zu erkennen als bei den Herkunftsnetzwerken. Zur Analyse werden die Ziel-Cluster mit den Herkunfts-Clustern verglichen. Zusätzlich wird pro Cluster über mögliche Ursachen für das Antwortverhalten aufgezeigt.

- **Cluster 0 - Address/Time - 4396**. Dies ist ein Cluster, der durch Firewalling geprägt ist. Er entspricht Cluster 1 der Herkunftsnetzwerke. ARIN ist in diesem Cluster am stärksten ausgeprägt. Für ein Zehntel dieses Clusters

werden auch Time-Exceeded-Nachrichten erhalten. Dies lässt auf den nicht geringen Anteil an RIPE-NCC und LACNIC-Netzwerken schließen.

- **Cluster 1 - Time/Echo - 3278**. Anfragen an diese Ziele führten zu Routing-Loops. Dies ist der einzige Cluster, wo der Anteil an RIPE-Netzwerken von einem der anderen RIRs übertroffen wird. Mit 42% der betroffenen Netzwerke in diesem Cluster trägt LACNIC am meisten zu den Routing-Loops bei. Der Anteil der anderen RIRs sollte jedoch auch beachtet werden. Der Anteil an ARIN-Netzwerken ist im Verhältnis zu den anderen Clustern geringer. Dies gilt auch für RIPE-Netzwerke, sie machen jedoch noch immer 29% der Netzwerke in diesem Cluster aus. Der Anteil an APNIC-Netzwerken ist im Vergleich zu Cluster 0 gestiegen. AFRINIC trägt mit weniger als 1% nur geringfügig zu Routing-Loops bei.

- **Cluster 2 - No-Route - 2187**. Für Netzwerke aus diesem Cluster werden oftmals nur No-Routes erhalten. Die Ursache wird wie bei den Herkunftsnetzwerken (Cluster 3) bei der Routen-Aggregierung gesehen. Die in den BGP Daten aggregierten Einträge führen bei dem Subnetzscan zu zahlreichen inaktiven Zielen. RIPE-NCC sowie LACNIC scheinen dieses Verhalten durch zahlreiche Netzwerke, die mit No-Route antworten, zu bestätigen. Der Anteil an ARIN, APNIC und AFRINIC Netzwerken ist verhältnismäßig gering. Die geringere Zahl an Netzwerke von ARIN kann durch die höheren Anteile in Firewalling-Clustern ausgelöst werden, da dadurch No-Route-Antworten häufiger unterdrückt werden.

- **Cluster 3 - Reject/Echo - 1336**. Ein Cluster, dessen Netzwerke sehr restriktiv konfiguriert sind und für deren Subnetze Reject-Routes definiert wurden. Die Echo-Replies sprechen für explizit konfigurierte Reject-Routen und nicht für ACLs, die die gesamten eingehenden Pakete blocken. Der Anteil an LACNIC-Netzwerken ist im Vergleich zu Cluster 2 und 3 wieder stark zurückgegangen. Die Anteile von RIPE-NCC, ARIN, APNIC und AFRINIC sind jedoch gestiegen.

- **Cluster 4 - Echo - 984**. Ein Cluster mit Netzwerken die kaum Error-Messages zurückschicken. Für diese Netzwerke wurden fast ausschließlich Echo-Replies erhalten. Dies stimmt mit Cluster 0 der Herkunftsnetzwerke überein. Bei diesen wurde ebenfalls der gegenüber den meisten anderen Clustern gestiegene Anteil an RIPE-NCC Netzwerken festgehalten.

- **Cluster 5 - Admin - 339**. Ein weiterer von Firewalling betroffener Cluster ist Cluster 5, wobei Administratively-Prohibited gemäß RFC der richtige Antworttyp für diesen Sachverhalt ist. In diesem Cluster ist wie in Cluster 0 ein deutlicher Anstieg des Anteils an ARIN-Netzwerken bemerkbar. Der Anteil an LACNIC-Netzwerken geht im Vergleich zu anderen Clustern stark zurück. Der Anteil an RIPE, APNIC und AFRINIC Netzwerken ist durchschnittlich.

- **Cluster 6 - Address/Echo - 895**. Ein Cluster mit sehr offen konfigurierten Netzwerken. Für Netzwerke aus

diesem Cluster werden sowohl Echo-Replies als auch Address-Unreachable Antworten erhalten. Die Kombination aus Address-Unreachable und Echo-Replies deutet auf den Einsatz von Konfigurationen ohne Firewalls hin. Das macht diese Netzwerke für weitere Scans sehr interessant. Der Anteil an RIPE-NCC Netzwerken ist am größten, es sind jedoch auch die Anteile von LACNIC, ARIN und AFRINIC Netzwerken durchschnittlich gegenüber den anderen Clustern. Nur der Anteil an APNIC-Netzwerken ist verhältnismäßig gering.

6 Fazit

Im Rahmen dieser Arbeit wurde eine internetweite ICMPv6-Messung durchgeführt. ICMP ist sowohl im IPv4- als auch im IPv6-Internet eines der fundamentalen Protokolle. Definiert werden diese Protokolle in sogenannten Request-for-Comments, welche Implementierungsrichtlinien für Protokolle im Internet darstellen. In RFC4443 wird ICMPv6 definiert. Der RFC beinhaltet zum Beispiel, wann welcher ICMPv6-Antworttyp zum Einsatz kommen soll. Wie diese Richtlinien umgesetzt werden, kann durch Internet-weite Messungen überprüft werden. Kapitel 2 legt zunächst dar, wie sehr sich der IPv6-Adressraum von IPv4 unterscheidet. Im Gegensatz zu IPv4 kann in IPv6 nicht mehr an jede mögliche Adresse ein Paket geschickt werden. Kapitel 3 beschreibt die Rahmenbedingungen der Messung. Mittels eines eigens adaptierten Tools wurden Echo-Requests an alle gerouteten /32-Präfixe und deren Subnetze versandt. Dabei wurden in Summe über 140 Milliarden Anfragen ins IPv6-Internet geschickt. In Kapitel 4 werden das Tool und für die Messung gemachte Anpassungen beschrieben. Es wird erläutert, warum trotz dieser enormen Anzahl an Paketen keine Zielnetzwerke überlastet werden. Messungen dieser Art laufen oftmals über Zeitspannen von mehreren Wochen. Dadurch entsteht der Bedarf, Serverwartungen auch während laufenden Messungen durchführen zu können. Es wird erstmalig gezeigt, wie der Highspeed-Scanner-ZMAP unterbrochen werden kann, um Scans zu einem späteren Zeitpunkt fortzusetzen. Eine Messung dieser Größenordnung führt auch zu einer dementsprechenden Anzahl an Antworten, welche gespeichert werden müssen. Es wurde über Vor- und Nachteile von Logging im Binärformat diskutiert und sich für diese Messung bewusst für das Speichern der Information im Textformat entschieden. Es wurde festgestellt, dass dieses nur geringfügig mehr Speicherplatz verbraucht, aber den Vorteil bietet, dass die Ergebnisse direkt in lesbarer Form vorliegen. In Kapitel 6 wurden die Messergebnisse schließlich ausgewertet. Um zu vergleichen, welchen Einfluss unterschiedliche Scangeschwindigkeiten auf die Ergebnisse haben, wurden Messungen mit 5 verschiedenen Scan-Raten durchgeführt. Durch die Auswertung wird deutlich, dass zwar die Zahl der Antworten mit steigender Scan-Rate exponentiell abnimmt, jedoch die Anzahl der Adressen, antwortenden Netzwerke und Zielnetzwerke, für welche eine Antwort erhalten wurde, annähernd gleich bleibt. Daher wurde der Einfluss der Scan-Raten auf die einzelnen Antworttypen ausgewertet. Es wird erstmalig gezeigt, dass sich Scan-Raten unterschiedlich auf die einzelnen Antworttypen auswirken. Die Antworttypen Reject-Route und Address-Unreachable sind besonders stark betroffen. Es gilt jedoch generell, dass der Einfluss von Rate-Limiting geringer ist, je mehr antwortende Adressen ein Antworttyp aufweist. Von dieser Beobachtung profitieren zum Beispiel die Antworttypen Administratively-Prohibited

und No-Route, da sie von einer verhältnismäßig hohen Anzahl an Adressen stammen. Die Messung konnte jedoch nicht alle 2^{32} möglichen Subnetzwerke pro Netzwerk abdecken, deswegen wurden diese auf 86000 pro Netzwerk reduziert. Pro Scan-Rate wurden zehn Messungen mit unterschiedlichen Subnetzwerken als Ziel ausgeführt. Aufgrund mangelnder Unterschiede der bei den Scan-Raten verglichenen Kategorien erfolgen weitere Auswertungen nur noch auf Basis einer dieser zehn Messungen.

Beim Vergleich der Scan-Raten ist aufgefallen, dass die Anzahl der Echo-Replies alle Erwartungen übertreffen. Ursache für dieses Phänomen ist Aliasing. Dabei antwortet ein Host aus dem Zielnetzwerk für alle anderen Adressen aus dem Netzwerk mit Echo-Replies. Es macht dabei keinen Unterschied, ob die ursprüngliche Zieladresse aktiv oder nicht aktiv ist. Dadurch verlieren Echo-Replies ihren Informationsgehalt. Schlimmer, es geht auch der Informationsgehalt für Echo-Replies aus anderen Netzwerken verloren, wenn Netzwerke mit Aliasing nicht aussortiert werden. In dieser Arbeit wird ein neues Verfahren zum Aussortieren von Netzwerken mit Aliasing vorgestellt. Im Vergleich zu anderen Verfahren zur Alias-Erkennung konnte das neue Verfahren direkt auf die Scan-Ergebnisse angewandt werden und es musste kein extra Scan zur Alias-Erkennung erfolgen. Mittels des Verfahrens wurden 278 Netzwerke als *aliased* klassifiziert, welche für mehr als 99% der Echo-Replies verantwortlich sind. Durch eine Zuordnung dieser Netzwerke zu den RIRs wurde festgestellt, dass ein Großteil dieser Netzwerke zu den Adressbereichen der RIPE-NCC gehören. Neben Aliasing wurden einige Netzwerkkonfigurationen erkannt, die eine überdurchschnittlich hohe Anzahl an antwortenden Adressen zeigen. Es wird daher eine 2-Sigma-Filterung angewandt, um solche Konfigurationen auszusortieren. Zu den Ausreißern zählen bei maximaler Scan-Rate 30 Netzwerke, welche für 88% der No-Route und 92% der Administratively-Prohibited Herkunftsadressen verantwortlich sind. Anschließend wurde der Beitrag der Antworttypen zu den Kategorien Antworten, Adressen, antwortende Netzwerke und antwortende Subnetzwerke ausgewertet. Die Antworttypen unterscheiden sich je nach Kategorie sehr stark. Je nachdem, auf welche Kategorien in einem Scan Wert gelegt wird, sollte daher auf unterschiedliche Antworttypen geachtet werden. In den Kategorie Adressen tragen No-Route, Administratively-Prohibited und Address-Unreachable besonders bei. Zur Erkennung von aktiven Netzwerken sollte auf Echo-Replies und Address-Unreachable geachtet werden. Anschließend wurde ausgewertet, woher die Antworten stammen. Bei maximaler Scan-Rate stammte nur ein Zehntel der Antworten aus dem Zielnetzwerk. Durch Auswerten der gleichen Bits zwischen Herkunfts- und Zieladresse konnte die ungefähre Herkunft der Antworten bestimmt werden, welche nicht aus dem Zielnetzwerk stammen. Rund 50% der Antworten unterscheiden sich bereits in den ersten 4-8 Bits. Bei diesen 50% gehören somit Herkunfts- und Zieladresse zu unterschiedlichen RIRs.

Abschließend wird überprüft, ob Herkunfts- und Zielnetzwerke sich nach Antwortverhalten in Gruppen einteilen lassen. Zur Clusteranalyse wird k-means eingesetzt. Die Ergebnisse zeigen sowohl bei Herkunfts- als auch bei Zielnetzwerken Cluster mit sehr stark ausgeprägten Merkmalen. Zur Auswertung der Antwortverhalten erfolgte eine Aufteilung in in sieben Cluster. Je nach Mittelwert der Antworttypen wurden mögliche Ursachen für diese Cluster diskutiert. Zu den Ursachen für gebildete Cluster zählen stark von Firewalling betroffene Netzwerke als auch Netzwerke, die

Routing-Loops aufweisen, oder zum Teil sehr offen konfigurierte Netzwerke. Kombiniert mit der Zugehörigkeit zu den RIRs wurde ein verstärktes Aufkommen von Routing Loops in Adressbereichen der LACNIC festgestellt. ARIN- und APNIC-Netzwerke sind stärker von Firewalling betroffen als andere RIRs. RIPE-NCC-Netzwerke sind in sehr offen konfigurierten Netzwerken stärker vertreten.

Dies hat zahlreiche Konsequenzen für den IPv6 Adressraum. Es gibt sehr starke geographische Unterschiede. Es wurden zahlreiche Routing Loops in IPv6 festgestellt, welche gemäß RFC4443 nicht vorhanden sein dürften. Netzwerkbetreiber sollten auf diese Fehlkonfigurationen aufmerksam gemacht werden, um unnötige Last vom Netzwerk zu nehmen. Die Vorgaben in RFC4443 sollten besser beachtet werden. Der Antworttyp Address-Unreachable, der eingesetzt werden kann, um Routing-Loops zu verhindern, wurde vor allem für Firewalling benutzt. Dies zerstört den eigentlichen Informationsgehalt von Adress-Unreachable Antworten. Durch das Clustern wurde jedoch auch eine Gruppe an Netzwerken erkannt, die neben Address-Unreachable auch Echo-Replies zurückschicken. Bei diesen Netzwerken ist es möglich durch Address-Unreachable-Antworten aktive Subnetze abzuleiten.

Diese Messung bietet Grundlage für zahlreiche zukünftige Arbeiten. Es sollte der Ursprung von Routing Loops in IPv6 genauer untersucht werden. Die Arbeit bietet die Möglichkeit, nach Netzwerken zu filtern, welche es ermöglichen, ICMPv6-Address-Unreachable auszunutzen, um aktive Subnetze zu erkennen. Mit dieser Information können neue, bessere Scantools erstellt werden. Die Entwicklung der Cluster sollte auch in Zukunft beobachtet werden, denn nur so kann kontrolliert werden, ob sich die Anzahl der Fehlkonfigurationen verringert und eine reibungslose Umstellung des Internets auf IPv6 garantiert werden.

Abbildungsverzeichnis

2.1 Vermittlungstypen (links: Leitungsvermittlung - rechts: Paketvermittlung) 6
2.2 Ursprung Antworttypen 8
2.3 ICMPv6-Error-Message-Header [9] 9
2.4 Visualusierung des Adressraums durch die Universität Oregon im Jänner 2019 [20]. 11
2.5 Regional-Internet-Registry Regionen[23] 13
2.6 Aggregierbare Global-Unicast-Adressen [26, p.7] 13
2.7 Global-Unicast-Adressformat Neu [26, p.7] 14
2.8 Global-Unicast-Adressformat EUI-64 [26, p.7] 14

4.1 Sendealgorithmus für Subnetzscan in ZMAP [6] 24
4.2 Sharding von ZMAP mit drei Threads (=Shards) [45] 25
4.3 Variablen für Interrupt 26
4.4 Output Varaiblen 26

5.1 Ursprung Antworttypen 30
5.2 Einfluss von Scan-Raten auf Ergebnisse 31
5.3 Antworten in Bezug auf Scan-Rate 32
5.4 Adressen in Bezug auf Scan-Rate 33
5.5 Präfixe mit Aliasing in Bezug auf RIRs 37
5.6 Bit-Vergleich Antworten 41
5.7 Datenformat für Antwortverhalten 44
5.8 Binäres Datenformat für Antwortverhalten 46
5.9 Antworten aus nicht allokierten Adressbereichen 50
5.10 Herkunftsnetzwerke-Cluster in Bezug auf RIRs 51
5.11 Zielnetzwerke Cluster in Bezug auf RIRs 56

Tabellenverzeichnis

2.1 ICMPv6 Typen [9] . . . 7
2.2 Vergleich IPv6-Global-Unicast-Adressraum mit und ohne HD-Ratio (Einheiten: 1B=1E9, 1T=1E12, 1Q=1E15, 1QQ=1E18) [31] . . . 15
2.3 ISP mit /32-Allokation und Vergabegröße von /48 – HD-Ratios [35] . . . 15
2.4 Scan-Tools und ihre Einsatzgebiete . . . 17

3.1 Zuordnung der BGP-Präfixe zu IANA vergebenen IPv6-Blöcken . . . 20
3.2 /32-Präfixe und zugehörige RIRs . . . 20
3.3 /32-Input-Präfixe für Scan und zugehörige RIRs . . . 21

4.1 ICMPv6 Klassifizierung in ZMAP . . . 27

5.1 Scanergebnisse zehn Seeds zu fünf Scan-Raten . . . 30
5.2 Scan-Ergebnisse je nach Antworttyp . . . 35
5.3 Ergebnisse Alias-Erkennung . . . 36
5.4 Ergebnisse 2-Sigma Regel . . . 38
5.5 Beitrag der Antworttypen zu den Scan-Ergebnissen . . . 39
5.6 Verteilung von IPv6-Blöcken durch IANA 2006 (4 bis 8 gleiche Bits) [22] . . . 42
5.7 Verteilung von IPv6-Blöcken durch IANA (8 bis 12 / 12 bis 16 gleiche Bits) [22] . . . 42
5.8 ICMPv6-Fehlernachrichten aus dem Zielnetzwerk . . . 43
5.9 Cluster k=7 Herkunftsnetzwerke . . . 45
5.10 Cluster 0 *Centroids* mit k=7 . . . 45
5.11 Cluster k=4 Herkunftsnetzwerke . . . 46
5.12 Cluster *Centroids* mit k=4 . . . 47
5.13 Cluster k=7 Herkunftsnetzwerke . . . 47
5.14 Cluster *Centroids* mit k=7 . . . 48
5.15 Antwortende Netzwerke in Bezug auf IANA-Allokationen . . . 50
5.16 Cluster k=7 Zielnetzwerke . . . 53

5.17 Cluster *Centroids* mit k=7 54
5.18 Herkunft Zielnetzwerke 55

Literaturverzeichnis

[1] O. Gasser, Q. Scheitle, S. Gebhard, and G. Carle, "Scanning the ipv6 internet: Towards a comprehensive hitlist," in *Proc. of 8th Int. Workshop on Traffic Monitoring and Analysis*, Louvain-la-Neuve, Belgium, Apr. 2016.

[2] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczynski, S. D. Strowes, L. Hendriks, and G. Carle, "Clusters in the expanse: Understanding and unbiasing ipv6 hitlists," in *Proceedings of the 2018 Internet Measurement Conference*. New York, NY, USA: ACM, 2018.

[3] J. Ullrich, P. Kieseberg, K. Krombholz, and E. Weippl, "On reconnaissance with ipv6: A pattern-based scanning approach," in *2015 10th International Conference on Availability, Reliability and Security*, Aug 2015, pp. 186–192.

[4] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson, "Target generation for internet-wide ipv6 scanning," in *Proceedings of the 2017 Internet Measurement Conference*. ACM, 2017, pp. 242–253.

[5] R. Beverly, R. Durairajan, D. Plonka, and J. P. Rohrer, "In the ip of the beholder: Strategies for active ipv6 topology discovery," in *Proceedings of the Internet Measurement Conference 2018*. ACM, 2018, pp. 308–321.

[6] F. Holzbauer, "Internet-wide scanning of ipv6 using custom version of zmap," 2018, bachelor's thesis.

[7] F. Gont and T. Chown, "Network reconnaissance in ipv6 networks," Internet Requests for Comments, RFC Editor, RFC 7707, March 2016.

[8] S. Bano, P. Richter, M. Javed, S. Sundaresan, Z. Durumeric, S. J. Murdoch, R. Mortier, and V. Paxson, "Scanning the internet for liveness," vol. 48, no. 2. ACM, 2018, pp. 2–9.

[9] A. Conta, S. Deering, and M. Gupta, "Internet control message protocol (icmpv6) for the internet protocol version 6 (ipv6) specification," Internet Requests for Comments, RFC Editor, RFC 4443, March 2006, http://www.rfc-editor.org/rfc/rfc4443.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4443.txt

[10] J. Postel, "Internet control message protocol," Internet Requests for Comments, RFC Editor, STD 5, September 1981, http://www.rfc-editor.org/rfc/rfc792.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc792.txt

[11] C. Franzetti, *Netzwerke*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2019, pp. 87–99. [Online]. Available: https://doi.org/10.1007/978-3-662-58534-4_8

[12] V. G. Cerf and R. E. Icahn, "A protocol for packet network intercommunication," *ACM SIGCOMM Computer Communication Review*, vol. 35, no. 2, pp. 71–82, 2005.

[13] Iljitsch and Beijnum, "Ggp, egp and 25 years of bgp: a brief history of internet routing," 2015. [Online]. Available: https://www.routerfreak.com/ggp-egp-and-25-years-of-bgp-a-brief-history-of-internet-routing/ (last access April 28, 2020).

[14] E. Davies and J. Mohacsi, "Recommendations for filtering icmpv6 messages in firewalls," Internet Requests for Comments, RFC Editor, RFC 4890, May 2007, http://www.rfc-editor.org/rfc/rfc4890.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4890.txt

[15] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications," in *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, 2013, pp. 605–620.

[16] V. Cerf, Y. Dalal, and C. Sunshine, "Specification of internet transmission control program," Internet Requests for Comments, RFC Editor, RFC 675, December 1974, http://www.rfc-editor.org/rfc/rfc675.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc675.txt

[17] J. Postel, "Internet protocol," Internet Requests for Comments, RFC Editor, STD 5, September 1981, http://www.rfc-editor.org/rfc/rfc791.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc791.txt

[18] S. E. Deering and R. M. Hinden, "Internet protocol, version 6 (ipv6) specification," Internet Requests for Comments, RFC Editor, RFC 1883, December 1995, http://www.rfc-editor.org/rfc/rfc1883.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc1883.txt

[19] V. Cerf, "I remember iana," Internet Requests for Comments, RFC Editor, RFC 2468, October 1998.

[20] R. Finnie, "Visual representations of the Internets in-use IPv4 and IPv6 address spaces," https://vad.solutions/ipmap/, 2019.

[21] V. Fuller and T. Li, "Classless inter-domain routing (cidr): The internet address assignment and aggregation plan," Internet Requests for Comments, RFC Editor, BCP 122, August 2006, http://www.rfc-editor.org/rfc/rfc4632.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc4632.txt

[22] "Ipv6 global unicast address assignments," 2018. [Online]. Available: https://www.iana.org/assignments/ipv6-unicast-address-assignments/ipv6-unicast-address-assignments.xhtml (last access March 11, 2019).

[23] "Ripe ncc service region," Tech. Rep., 2018. [Online]. Available: https://www.ripe.net/about-us/what-we-do/ripe-ncc-service-region (last access June 2, 2020).

[24] "Ipv6 address allocation and assignment policy," Tech. Rep., 2020. [Online]. Available: https://www.ripe.net/publications/docs/ripe-738 (last access March 22, 2020).

[25] T. Narten, G. Huston, and L. Roberts, "Ipv6 address assignment to end sites," Internet Requests for Comments, RFC Editor, BCP 157, March 2011.

[26] R. M. Hinden and S. E. Deering, "Ip version 6 addressing architecture," Internet Requests for Comments, RFC Editor, RFC 2373, July 1998, http://www.rfc-editor.org/rfc/rfc2373.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc2373.txt

[27] R. Hinden, S. Deering, and E. Nordmark, "Ipv6 global unicast address format," Internet Requests for Comments, RFC Editor, RFC 3587, August 2003.

[28] C. Huitema, "The h ratio for address assignment efficiency," Internet Requests for Comments, RFC Editor, RFC 1715, November 1994, http://www.rfc-editor.org/rfc/rfc1715.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc1715.txt

[29] A. Durand and C. Huitema, "The h-density ratio for address assignment efficiency an update on the h ratio," Internet Requests for Comments, RFC Editor, RFC 3194, November 2001.

[30] RIPE, "Best current operational practice for operators: Ipv6 prefix assignment for end-users - persistent vs non-persistent, and what size to choose," https://www.ripe.net/publications/docs/ripe-690, 2017.

[31] A. Durand, "Analyzing IPv4 and Ipv6 address space with the HD-ratio," 2002. [Online]. Available: https://tools.ietf.org/html/draft-durand-hdv4v6-00

[32] "Population of the world and countries." [Online]. Available: https://countrymeters.info/en (last access May 29, 2020).

[33] A. A. Akplogan, "Policy to change the ipv6 hd ratio from 0.8 to 0.94 | afpub-2007-v6-002," https://afrinic.net/policy-to-change-the-ipv6-hd-ratio-from-0-8-to-0-94-afpub-2007-v6-002, 2007.

[34] LACNIC, "IPv6 Address Allocation and Assignment Policies," 2020.

[35] "Ipv6 address allocation and assignment policy," Tech. Rep., 2013. [Online]. Available: https://www.apnic.net/community/policy/ipv6-address-policy_obsolete/ (last access May 28, 2020).

[36] C. B. Lee, C. Roedel, and E. Silenok, "Detection and characterization of port scan attacks," *Univeristy of California, Department of Computer Science and Engineering*, 2003.

[37] G. Lyon, "Nmap security scanner," *Nmap. org,[En línea]. Available: http://nmap. org/.[Último acceso: 20 abril 2015]*, 2014.

[38] T. Fiebig, K. Borgolte, S. Hao, C. Kruegel, and G. Vigna, "Something from nothing (there): Collecting global ipv6 datasets from dns," in *Passive and Active Measurement*, M. A. Kaafar, S. Uhlig, and J. Amann, Eds. Cham: Springer International Publishing, 2017, pp. 30–43.

[39] K. Borgolte, S. Hao, T. Fiebig, and G. Vigna, "Enumerating active ipv6 hosts for large-scale security scans via dnssec-signed reverse zones," in *2018 IEEE Symposium on Security and Privacy (SP)*, May 2018, pp. 770–784.

[40] F. Gont, "scan6 - an ipv6 host scanner," 2013. [Online]. Available: http://manpages.ubuntu.com/manpages/cosmic/man1/scan6.1.html (last access March 11, 2019).

[41] C. Kukovic, "Ipv6 high performance scanning," TU Wien, Tech. Rep., 2016. [Online]. Available: http://repositum.tuwien.ac.at/obvutwhs/content/titleinfo/1553607

[42] "Archipelago (ark) measurement infrastructure," 2007. [Online]. Available: http://www.caida.org/projects/ark/ (last access March 5, 2019).

[43] E. W. Gaston, "High-frequency mapping of the ipv6 internet using yarrp," Naval Postgraduate School Monterey United States, Tech. Rep., 2017.

[44] "Pf ring zc (zero copy)." [Online]. Available: https://www.ntop.org/products/packet-capture/pf_ring/pf_ring-zc-zero-copy/ (last access March 21, 2019).

[45] D. Adrian, Z. Durumeric, G. Singh, and J. A. Halderman, "Zippier zmap: internet-wide scanning at 10 gbps," in *8th {USENIX} Workshop on Offensive Technologies ({WOOT} 14)*, 2014.

[46] M. Wiedenbeck and C. Züll, "Klassifikation mit clusteranalyse: Grundlegende techniken hierarchischer und k-means-verfahren," 2001.

[47] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," Stanford, Tech. Rep., 2006.

[48] "k-means." [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/k_means.html (last access May 22, 2020).