Why the Internet Sucks: A Core Perspective

Yves Müller
Freie Universität Berlin
Berlin, Deutschland
uves@spline.de

Zusammenfassung

Die vorhandene Routing-Struktur des Internets erweist sich zunehmend als unzulänglich. Eines der Hauptprobleme stellt die hohe Belastung der Router in der Default-Free-Zone dar, die unter anderem verursacht wird durch die Nutzung von Multihoming in vielen Teilnetzwerken. Die Lösung dieses und weiterer Probleme erfordert grundsätzliche Änderungen an der vorhandenen Struktur. Dabei bieten sich zwei prinzipielle Lösungsansätze: die Separation und die Elimination. Diese Arbeit soll die Ursachen untersuchen und die Löungsmöglichkeiten vergleichen, um mögliche Perspektiven für die zukünftige Struktus der Internets aufzuzeigen.

Inhaltsverzeichnis

1 Einleitung

Durch die starke Zunahme an Teilnehmern in den letzten Jahren erreicht das Internet seine technischen Grenzen auf der Netzwerkschicht. Probleme zeigen sich unter anderem bei der Anzahl an verfügbaren Adressen, der Mobilität von Endgeräten, der Sicherheit des Gesamtsystems und den Routing-Protokollen. Durch IPv6 wurde für die Anzahl der Adressen und Teilaspekte der Mobilität eine Lösung gefunden. Auch das Thema Sicherheit, als begleitender Prozess jeder Entwicklung, gewann in den letzten Jahren mehr Aufmerksamkeit. Der Handlungsbedarf beim Internet-Routing wurde schon vor einiger Zeit erkannt [1]. Der hohe Verteilungsgrad, die hohe Anzahl an Akteuren und die Einflüsse auf andere Subsysteme des Internets erschwerten die Suche nach neuen Lösungen für das Routing-System. Trotzdem gibt es interessante Lösungsvorschläge, um auch diesen Grundbestandteil des Internets auf zukünftige Anforderungen vorzubereiten.

Während der letzten Jahre wuchs die Größe der Routing-Tabelle in der Default-Free-Zone (DFZ) exponentiell an, trotz nur linearem Wachstum der verfügbaren Adressen[8]. Da das Routing-System mit wachsender Netzwerkgröße nicht skaliert, ist seine Funktion gefährdet [10]. Einen der größten Faktoren für diesen Anstieg stellt redundante Anbindung von Edge-Netzwerken an das Internet (Multihoming) dar. Neben der steigenden Tabellengröße gefährdet auch die daraus resultierende hohe Frequenz an Updates das Routing-System. Alle bisherigen Lösungsvorschläge fallen in zwei grundsätzliche Kategorien: Separation oder Elimination [10].

Diese Arbeit soll aufzeigen, welche aktuellen Entwicklungen die Skalierbarkeit des Internet-Routing gefährden und die Gründe dafür untersuchen. Es werden Lösungsansätze vorgestellt, die zur Zeit innerhalb von Forschung und Standardisierung diskutiert werden.

Der erste Abschnitt soll in das Thema einleiten und die nötigen Begriffe erläutern, um anschließend Probleme sowie deren Ursachen und Auswirkungen genauer zu untersuchen. Zwei prinzipielle Lösungsmöglichkeiten werden im dritten Abschnitt aufgezeigt, wobei Ansätze zur Separation ausführlich diskutiert werden. Schlussendlich werden zuvor gewonnene Erkenntnisse zusammengefasst und bewertet.

2 Problemstellung

2.1 Grundlagen

Mit Routing wird der Vorgang bezeichnet, der nötig ist, um ein Datenpaket in einem Netzwerk weiterzuleiten [11]. Auf dem Weg des Paketes von

der Quelle zum Ziel muss an jedem Router entschieden werden, zu welchem seiner direkten Nachbarn (nächster Knoten auf der Netzwerkschicht) ein Paket weitergeleitet werden soll. Diese Entscheidung wird anhand der Routing-Tabelle getroffen. Eine Routing-Tabelle enthält Zuordnungen zwischen Präfixen und Nachbarn, über die man den Präfix erreichen kann [11]. Erhält der Router ein Paket, so geht er die Routing-Tabelle durch und ermittelt die Menge der Präfixe in denen die Zieladresse enthalten ist [15]. Anschließend leitet er das Paket an den Nachbarn weiter, der dem längsten dieser Präfixe zugeordnet ist. Die Einträge in der Routing-Tabelle werden entweder statisch festgelegt oder durch ein Routing-Protokoll ausgetauscht [11]. Routing-Protokolle werden unterschieden in Interior Gateway Protokolle (IGP) und Exterior Gateway Protokolle (EGP) [15]. Mit EGPs werden Routing-Informationen zwischen autonomen Systemen ausgetauscht, während IGPs innerhalb eines autonomen Systems verwendet werden.

Das Internet besteht aus kleinen, individuellen, selbständigen Netzwerken. Ein autonomes System (AS) wird von einer Menge solcher Netzwerke gebildet, die untereinander verbunden sind und eine gemeinsame EGP Routing-Policy besitzen [7]. Behandelt ein Netzwerk ausschließlich Pakete, deren Quelle oder Ziel es selbst ist, wird es als Edge-Netzwerk (EN) bezeichnet. Ein Internet Service Provider ist ein AS, das Verkehr an Edge-Netzwerke oder andere autonome Systeme weiterleitet. Wenn ein ISP keine Edge-Netzwerke bedient, handelt es sich um ein Transit-AS. Alle autonomen Systeme, die keine Pakete für andere weiterleiten, werden als Stub-AS bezeichnet. Sie werden, je nach der Anzahl ihrer Verbindungen zu ISPs, in die Kategorien singelhomed und multihomed eingeteilt [11]. Ein Edge-Netzwerk kann ein eigenständiges Stub-AS bilden oder Teil des autonomen Systems seines Providers sein. Die Bildung eines autonomen Systems aus einem oder mehreren Netzwerken ist insbesondere dann nötig, wenn Routing-Informationen mit anderen autonomen Systemen ausgetauscht werden sollen [7].

Eine strikt hierarchische Vergabe von Präfixen führt dazu, dass die Komplexität des Routings durch die Anzahl der Subpräfixe gesteuert werden kann. Jeder Router muss ermitteln, ob das Ziel in einem seiner Subpräfixe liegt, anderenfalls kann er das Paket entlang einer Default-Route an den übergeordneten Router weiterleiten. Der Router an der Wurzel dieser Baumstruktur verfügt über keine Default-Route und muss daher Pakete zu unbekannten Präfixen verwerfen. Eine solche Hierarchie ist jedoch für das Internet nicht wünschenswert, denn sie erfordert eine zentrale Instanz als Wurzel mehrerer Teilbäume. Dieser Knoten kann hohen Lasten ausgesetzt sein und bildet eine kritische Fehlerquelle für seinen gesamten Unterbaum. Das Internet bildet eine flache Struktur auf AS-Ebene, wobei die einzelnen

Komponenten wieder hierarchisch aufgebaut sind, um die Komplexität des Routing-Systems zu begrenzen. Das Konzept der Adressvergabe im Internet sieht vor, dass die RIRs den großen ISPs Präfixe zuteilen und diese anschließend die Adressen hierarchisch an ihre Kunden (kleine ISPs und Edge-Netzwerke) weitergeben [6]. In der Routing-Perspektive wird der vorhandene Adressraum unterteilt in die den IPSs zugeordneten Präfixe.

Es gibt Knoten im Internet, welche die gleiche Sicht auf alle nichthierarchischen Präfixe benötigen. Diese Knoten sind normalerweise die EGP-Router der ISPs. Sie besitzen keine Default-Routen und bilden daher die Default-Free-Zone (DFZ). Alle Knoten der DFZ benötigen eine Route zu jedem in der DFZ bekannten Präfix. Die Router eines Stub-AS können eine Default-Route zum Internet über ihre Provider verwenden und gehören dann nicht zur DFZ. Ihr Präfix wird trotzdem in der gesamten DFZ bekanntgegeben.

2.2 Kernproblem

Die Skalierbarkeit des Routings zwischen autonomen Systemen ist vor allem dadurch gefährdet, dass die Edge-Netzwerke in den Routing-Tabellen der DFZ-Router abgebildet werden. Ein Edge-Netzwerk kann im Präfix seines Providers aggregiert sein oder eine eigene Route besitzen. Um letzteres zu ermöglichen nutzt das Edge-Netzwerk providerunabhängige Adressen (PI) oder deaggregierte Adressen seines Providers [10]. Ein providerunabhängiger Präfix wird direkt von der RIR vergeben und liegt nicht innerhalb des Präfix eines Providers. Ist der Präfix eines Edge-Netzwerkes nicht von seinem Provider aggregiert, muss der Pfad zum EN über seinen Provider in allen Routing-Tabellen der DFZ enthalten sein.

Providerunabhängige Adressen werden von Edge-Netzwerken vor allem aus zwei Gründen genutzt: Sie erleichtern den Wechsel des Providers und sie bieten eine für das Edge-Netzwerk einfache Möglichkeit Multihoming zu realisieren. Nutzt ein Kunde den Subpräfix seines Providers, so erhält er beim Wechsel des ISPs einen anderen Präfix. [10]. Dies erfordert die Änderung der IP-Adresse an jedem Gerät und ist dadurch eine aufwendige Operation. Viele Edge-Netzwerke nutzen Multihoming, um ihre Netzwerklast auf mehrere Provider zu verteilen oder im Falle eines Verbindungsausfalls auf einen anderen Provider zurückzugreifen. Um über alle seine angebundenen Provider adresstransparent erreichbar zu sein, muss der Präfix des ENs in der gesamten DFZ sichtbar sein. Kein ISP kann ein multihomed Edge-Netzwerk in seinen eigenen Präfix eingliedern [10], es sei denn es wird eine Struktur geschaffen, die diese Eingliederung für andere Endgeräte transparent macht.

2.3 Auswikungen und Faktoren

Bis zum Jahr 2004 wurden insgesamt ca. 63 000 IP-Adressblöcke registriert, davon ungefähr 18 000 in den letzten 7 Jahren. Im gleichen Zeitraum durchgeführte Messungen am meist verwendeten Exterior Gateway Protokoll BGP zeigten, dass die Routing-Tabellen der DFZ ca. 160 000 Einträge umfassten [13]. Dazu wurden die Routing-Tabellen verschiedener DFZ-Router aus ausgewählten autonomen Systemen zusammengeführt und ausgewertet. Die Anzahl der Einträge pro registriertem IP-Adressblock nahm von 1998 mit 1,33 Einträge auf 2,54 Einträge im Jahr 2004 zu [13]. Diese hohe Menge an Routen je Adressblock deutet darauf hin, dass viele Edge-Netzwerke Multihoming nutzen [8].

Ein weiterer Wachstumsfaktor für die Routing-Tabellen ist die unsaubere Allokation von Präfixen. Idealerweise sollte ein AS über genau einen Präfix verfügen, der alle Netzwerke des AS enthält [7]. Außerdem werden Präfixe, die eigentlich zusammengefasst werden können, über verschiedene autonome Systeme verteilt oder Subpräfixe werden deaggregiert, so dass die Anzahl der Routen in der DFZ steigt [13].

Mit der Größe der Routing-Tabelle wächst auch die Anzahl der Updates, die nötig sind, um die Tabelle zu aktualisieren. In der ersten Jahreshälfte 2004 wurden 24 000 Einträge in der BGP-Tabellen entfernt und 36 000 Einträge hinzugefügt, wobei sich die Anzahl der erreichbaren Adressen nur geringfügig veränderte [13]. Dies impliziert, dass nicht nur das Suchen in der Tabelle sondern auch das Pflegen der Tabelle aufwendiger wird und die DFZ-Router zusätzlich belastet.

Ein weiteres Problem ist, dass eine Unicast-Adresse in der Praxis eine überladene Bedeutung besitzt. Sie wird als Identifier genutzt, da sie einen Socket während eines Kommunikationsvorganges identifiziert. Gleichzeitig dient sie dem Routing-System als Locator. Ändert sich also der Präfix bezüglich eines Gerätes, erzwingt dies eine Unterbrechung der Kommunikation zwischen Sockets. Dies erschwert die Implementierung von Mobile-IP und Multihoming ohne in der Routing-Tabelle der DFZ sichtbar zu sein.

3 Lösungsansätze

3.1 Elimination

Ein Lösungsansatz ist die Elimination von providerunabhängigen Adressen, so dass alle Edge-Netzwerke einen Subpräfix aus dem Präfix ihres Providers nutzen. Nur der Präfix des Providers muss in der DFZ bekanntge-

geben werden, was zu einer kleineren und stabileren Routing-Tabelle führt, da es deutlich weniger Provider als ENs gibt und Provider nur geringen Veränderungen unterliegen.

Ohne providerunabhängige Adressen werden neue Ansätze für Multihoming benötigt. Das Edge-Netzwerk erhält von jedem seiner Provider einen Subpräfix, um durch jede Verbindung für die DFZ erreichbar zu sein. Die Endgeräte und von ihnen genutzte Protokolle müssen so erweitert werden, dass sie mit mehreren Adressen zur Paketübermittlung arbeiten können [10]. Ein Endgerät muss zum Paketversand möglichst alle Adressen seines Kommunikationspartners ermitteln und für Antworten alle eigenen Adressen mitteilen. Es muss erkennen, ob sich die Erreichbarkeit des Kommunikationspartners ändert und entsprechend reagieren. Die Nutzung von verschiedenen Adressen zur Paketübermittlung sollte für Anwendungen transparent sein. Um diese zu erreichen, sind Modifkationen an verschiedenen Diensten (z.B. DNS) und Protokollebenen (z.B. Transportschicht) denkbar. Einen möglichen Ansatz bietet shim6 [14], eine Erweiterung der Schnittstelle zwischen Transport- und Netzwerkschicht. Shim6 sorgt für den Austausch der möglichen Adressen, überprüft während der gesamten Kommunikation ihre Erreichbarkeit und schreibt ggF Adressen um, so dass ein Wechsel der Adressen auf beliebiger Seite für die Transportschicht transparent bleibt. Damit bietet shim6 auch die Möglichkeiten Mobile-IP zu betreiben, trotz der weiterhin bestehenden Doppelsymatik einer IP-Adresse.

Die Elimination von providerunabhängigen Adressen führt dazu, dass sich der Präfix eines Edge-Netzwerkes beim Wechseln des Providers ändert. Eine mögliche Lösung bietet IPv6 mit der "Stateless Address Autoconfiguration". Sie ermöglicht einen Präfixwechsel des gesamten Netzwerkes, ohne Änderungen an jedem Endgerät vorzunehmen [16]. Allerdings sind auch hier Erweiterungen notwendig, um laufende Kommunikationsvorgänge nicht zu unterbrechen. Für Edge-Netzwerke, die IPv4 nutzen, existiert zur Zeit keine vergleichbar zuverlässige Lösung.

Generell muss die Intelligenz der Endgeräte erhöht werden. Dies widerspricht zwar nicht dem Prinzip des Internets, erschwert aber die Wartung großer Netze und stellt insbesondere ein Problem für eingebette oder stark ausgelastete Systeme dar.

Die hohen Anforderungen an die Endgeräte bei der Elimination machen ihre Umsetzung für die Betreiber der Edge-Netzwerke unattraktiv [10]. Die Vorteile der Elimination liegen auf Seite der ISPs, welche den stabilen Betrieb ihrer Router durch die wachsende Routing-Tabelle in der DFZ gefährdet sehen. Jedoch ist für die Umsetzung eine aktive Beteiligung der Edge-

Netzwerke erforderlich, da sie ihre providerunabhängigen Präfixe aufgeben müssen [10]. Aufgrund dieses Mangels an Motivation ist zu erwarten, dass nur eine langsame Einführung der Elimination möglich ist. Erst eine hohe Anzahl an Edge-Netzwerken, die auf ihren PI-Präfix verzichtet, bringt die gewünschten Effekte.

Ein Vorteil des Eliminationsansatzes ist, dass keine neuen Strukturen geschaffen werden müssen. Jedoch sind die Änderungen an vorhandenen Strukturen nötig, die für die Edge-Netzwerke nicht transparent sind.

3.2 Separation

Die Grundidee der Separation ist die doppelte Bedeutung einer IP-Adresse aufzuheben, indem Identifier und Locator separat zugeordnet werden. Der Ansatz wird daher auch als Identifier-Locator-Split bezeichnet [1],[10]. Routing-Entscheidungen in der DFZ werden anhand von RLOCs (routing locators) getroffen. Ein Endgerät wird durch den EID (Endpoint Identifier) eindeutig und dauerhaft identifiziert. An der Schnittstelle zwischen Zielnetzwerk und DFZ muss zwischen EIDs und RLOCs umgewandelt werden. Um dies zu ermöglichen, muss eine Zuordnung zwischen EIDs und RLOCs geschaffen werden. Dazu wird ein Mapping-System verwendet, das Auskunft erteilt über welche RLOCs ein gegebener EID zu erreichen ist.

Das Mapping-System ist eine Datenbank, welche Paarungen aus EIDs und RLOCs enthält, wobei die EIDs als Schlüssel dienen. Für den Paketversand wird ein Locator des Empfängers benötigt, so dass ohne Mapping ein EN nicht von anderen ENs erreicht werden kann. Um eine möglichst hohe Erreichbarkeit zu gewährleisten, sollte das Mapping-System also redundant und sicher vor Manipulation sein. Um bei Multihoming die Nutzung der einzelnen Providerverbindungen für einkommende Pakete genau zu steuern, sollte das Mapping eine Gewichtung der RLOCs bieten [12]. Es müssen ebenfalls Mechanismen geboten werden die eine zeitnahe direkte Einflussnahme des Netzwerkbetreibers auf das Mapping erlauben. Die Zuordnung zwischen EIDs und RLOCs sollte schnell durchgeführt werden, um geringe Latenzen zu erhalten. Insbesondere Caching kann helfen die Anzahl der vermutlich recht aufwendigen Anfragen an das Maping-System zu reduzieren. Im Widerspruch dazu sollten Änderungen am Mapping einer EID möglichst schnell für alle Router in der DMZ sichtbar werden.

Vorteile Ein ISP kann ohne Einschränkungen die RLOCs für seine Kunden aus seinem RLOC-Präfix entnehmen. Dies führt dazu, dass die Routing-Tabelle im Core erheblich kleiner und stabiler wird, da es weniger ISPs als Edge-Netzwerke gibt und die Anzahl der ISPs sowie ihre Routen nur geringer Veränderungen unterliegen [10], [1]. Um ein Edge-Netzwerk durch

mehrere Provider anzubinden, also um Multihoming zu betreiben, müssen nur die RLOCs aller Provider im Mapping-System mit dem EID-Präfix des Edge-Netzwerk assoziiert werden[2]. Dies ermöglicht Multihoming ohne Auswirkungen auf das Routing in der DFZ zu haben. Ein Providerwechsel wirkt sich nur auf die Zuordnung im Mapping-System aus, nicht auf das Core-Routing.
Da sich EIDs und RLOCs durch IP-Adressen dargestellt werden können, ist es weder notwendig die Endgeräte noch die Router innerhalb der DFZ zu modifizieren [10]. Nur an der Schnitstelle zwischen DFZ und Edge-Netzwerken sind Änderungen nötig.
Die Trennung von DFZ und Edge-Netzwerken durch eine Schnittstelle schafft mehr Modularität. Dies kann genutzt werden, um die beiden Strukturen unabhängig voneinander zu modifizieren und zu verbessern.
Die Gewichtung der RLOCs im Mapping-System erlaubt eine genaue Steuerung des Verkehrsflusses über die verschiedenen Provider eines multihomed Edge-Netzwerkes [12]. Für weitere Anwendungen, wie etwa das Reagieren auf DoS-Attacken, kann das Mapping-System auch genutzt werden [10].
Der Locator-Identifier-Split ermöglicht, dass sich Endgeräte bewegen ohne das sich ihr Identifier ändert. Somit ist die Kommunikation zwischen Endgeräten von Topologieänderungen unbeeinflusst. Dies bedingt jedoch ein ausreichend schnelles Mapping-System, da ohne korrekte Zuordnung keine Daten zum Endgerät geleitet werden können.

Nachteile Die Umwandlung zwischen RLOCs und EIDs ist mit einem gewissem Aufwand verbunden, der aber im Vergleich zu den Einsparungen beim DFZ-Routing vernachlässigt werden kann. Zentraler Bestandteil und kritischer Faktor des Separationsansatzes ist ein Mapping-System zwischen EIDs und RLOCs, das alle schon genannten Anforderungen erfüllen muss.

3.2.1 Lösungsansätze zur Separation

Für die Umwandlung zwischen RLOCs und EIDs existieren zwei verschiedene Ansätze. Beim Tunneling-Verfahren werden die EID-Pakete aus dem EN an der Schnittstelle zur DFZ in ein RLOC-Paket zum Routen innerhalb der DFZ gekapselt [2]. Ein anderes Verfahren verfolgen Address-Rewriting-Protokolle. Hier werden die Adressen des Ursprungspaketes umgeschrieben.

Tunneling mit LISP LISP erfordert keine Modifikationen an den Endgeräten in den Edge-Netzwerken [2]. Sie arbeiten bei der Adressierung weiterhin mit IPv4- oder IPv6-Adressen. Auch die Namensauflösung und das Routing innerhalb von Edge-Netzwerken muss nicht verändert werden. Wird ein Paket zum ISP übermittelt, erreicht es beim Provider den ITR (Ingress Tunnel Router). Dieser ermitelt zum Ziel-EID die RLOCs mittels des Mapping-Systems. Anschließend packt er das IP-Paket in ein LISP-Paket.

Als Zieladress wird die zuvor ermittelte RLOC benutzt, Quelladresse ist die RLOC des ITRs. Dann wird das Paket durch die DFZ übermittelt. Es erreicht schließlich den ETR (Egress Tunnel Router), der die Ziel-RLOC besitzt. Dieser ist an das EN angeschlossen, in dem sich die Ziel-EID befindet. Er packt das Paket aus und sendet es an das EN.

Durch die Kapselung des Ursprungspaketes kann es dazu kommen, dass das entstehende Paket die Maximum Transfer Unit (MTU) einer Verbindung innerhalb der DFZ überschreitet. LISP definiert zwei verschiedene Verfahren, um dieses Problem zu lösen [2]. Eine statusfreie Lösung ist es, Pakete ab einer bestimmten Größe zu verwerfen und eine ICMP-Meldung zum Absender zu schicken. Die zweite Lösung sieht vor, dass der ITR zu allen gecachten RLOCs die maximale Paketgröße speichert, mit der er dorthin senden konnte ohne eine ICMP-Meldung zu erhalten. Dies stellt sicher das die MTU voll ausgenutzt wird.

Address-Rewriting mit Six/One Router Im Gegensatz zu LISP wird bei Six/One das Umwandeln zwischen EIDs und RLOCs von Routern innerhalb des Edge-Netzwerkes übernommen [17]. Das EN verfügt für jede Verbindung zu einem ISP über ein Six/One Router. Dieser ist für die Umwandelung zwischen EIDs und RLOCs zuständig. Unabhängig davon, ob das Zielnetzwerk Six/One unterstützt, werden die Quelladressen ausgehender und die Zieladressen eingehender Pakete umgewandelt. Handelt es sich bei der Gegenseite um ein EN mit Six/One Router, werden ebenfalls die jeweils anderen Adressen umgeschrieben. Die orginalen Adressen werden im Six/One Extension Header des Paketes gespeichert, so dass eine Rückübersetzung am Ziel möglich ist. Um die RLOC des Zielnetzwerkes zu ermitteln wird das Mapping-System befragt.

3.2.2 Lösungsansätze für das Mapping

Im Zusammenhang mit LISP wurden schon zahlreiche Mapping-Systeme spezifiziert. Ihre Verwendung ist aber auch zusammen mit anderen Separationsansätzen denkbar. Die verschiedenen Vorschläge unterscheiden sich in der Verteilungsart der Informationen [12]. Einfache Systeme wie LISP-NERD verteilen die Mapping-Informationen aktiv mittels des Push-Mechanismus. Dabei werden die kompletten Mapping-Informationen auf allen Routern vorgehalten und es wird ein Mechanismus geschaffen, der alle Änderungen an die vorhandenen Knoten verteilt. Bei anderen Protokollen, wie zum Beispiel LISP-DHT, ermitteln die Router nur die Mappings, welche sie tatsächlich benötigen. Es gibt keine Instanz, die eine komplette Zuordung besitzt. Diese Protokolle haben eine komplexere Struktur und skalieren nach ersten Messungen gut [9]. Es gibt auch Mapping-Systeme die eine Zwischenform nutzen,

indem nur die existierenden EIDs aktiv verteilt werden. Ein Beispiel hierfür bietet LISP-ALT.

LISP-DHT Ein mögliches Mapping-System, basierend auf einer modifzierten Chord-Hashtabelle, beschreibt LISP-DHT [12]. Chord nutzt für jeden Teilnehmer eine ChordID und bildet eine ringförmige Struktur, die nach den ChordIDs geordnet ist. Die Chord-Knoten müssen sich den nächst größeren und nächst kleineren Knoten merken, um die Tabelle zu erhalten und Anfragen weiterzuleiten. Jeder Knoten ist für alle ChordIDs zuständig, die kleiner oder gleich seiner eigenen, aber größer als die seines Vorgängers sind. In LISP-DHT wird die größte von dem Knoten verwaltete EID als ChordID genutzt. Um einen Chord-Knoten zu adressieren, müssen seine EID und RLOCs gespeichert werden.
Jeder Chord-Knoten besitzt eine sogenannte Finger-Tabelle, um Abfragen effizienter durchzuführen. Der $i$te Eintrag der Tabelle enthält den $m + 2^{i-1}$ten Knoten der Chord, wobei m die Nummer des Knoten ist, der die Finger-Tabelle speichert. Die Korrektheit der Zuordnung ist nicht nötig, um die Funktionalität der Chord zu gewährleisten. Jedoch können mittels der Finger-Tabelle Abfragen in O(log n) Schritten bearbeitet werden.

Um dem Chord-Ring beizutreten muss wenigstens ein Chord-Knoten bekannt sein. Von diesem ausgehend sucht der beitretende Knoten in der bestehenden Chord seinen Vorgänger und Nachfolger und initialisiert seine Finger-Tabelle. Nun ist er bereit Anfragen an die Chord zu stellen, um beliebige Mappings zu erfahren. Aber er kann noch keine eigenen Mappings in die Chord einfügen, denn dazu ist eine Authentifikation nötig. Ein Knoten, der Mappings für einen bestimmten EID-Präfix liefern möchte, muss über ein Zertifikat verfügen, das ihn dazu berechtigt. Solche Zertifikate werden von der zuständigen RIR ausgestellt. Die Nachbarn des neuen Knotens überprüfen so, ob eine Berechtigung vorliegt und anschließend den Knoten als neuen Nachbarn aufnehmen.

Um Redundanz zu gewährleisten, sollte das Mapping für einen EID-Präfix mehrfach in der Chord gespeichert werden. Ein Ansatz wäre das Spiegeln des Mappings auf dem jeweiligen Nachbarknoten. Dadurch wird es für den Anbieter des Mappings jedoch schwieriger seine EIDs zu kontrollieren. Um dies zu vermeiden, können Redundanzgruppen gebildet werden. Sie enthalten mehrere Server, die das gleiche Mapping anbieten. Eine solche Redundanzgruppe wird in der Vorgänger- und Nachfolgerrelation sowie in der Finger-Tabelle anstatt eines einzelnen Chord-Knotens verwendet. Innerhalb der Redundanzgruppe ist eine Gewichtung zwischen den Servern möglich.

LISP-DHT bietet ein redundantes und robustes Mapping-System, das mittels Finger-Tabellen einen effizienten Zugriff auf die Mapping-Daten erlaubt. Das Mapping-System an sich besitzt keine Cache-artigen Strukturen. Es bietet dem Besitzer der EIDs immer direkte Kontrolle über die Zuordnung.

4 Zusammenfassung und Bewertung

Im Rahmen der Internet Enigneering Task Force (IETF) wurden bereits die einzelden Lösungsansätze, die in dieser Arbeit angesprochen werden, ausgearbeitet und diskutiert. Aktuell gibt es sowohl zu LISP als auch zu shim6 aktive Arbeitsgruppen innerhalb der IETF [4].

Bei der Elimination existiert bereits eine gültige Spezifikation des Protokolls shim6 durch den RFC 5533 [14],[3]. LISP scheint zur Zeit der einzige diskutierte Separationsansatz zu sein. Zu anderen Standards, wie zum Beispiel GSE, Six/One, TRRP oder APT finden sich im Internet-Archiv der IETF keine veröffentlichten Entwürfe. Der Entwurf zu einem LISP-RFC [2] existiert zur Zeit in der fünften Revision. Er wurde jedoch noch nicht dem Zuständigen Area Director der IETF vorgelegt und auch ein Antrag auf Veröffentlichung existiert noch nicht [3]. Multicast-Erweiterungen für LISP und die Mapping-Systeme befinden sich in einem vergleichbaren oder früheren Stadium.

Eine Diskussion der Lösungsvorschläge findet innerhalb der Routing Reasearch Group (RRG) statt, die Teil der Internet Resarch Task Force ist [5]. Ein mögliches Ergebnis der RRG könnte die Empfehlung einer Lösung an die IETF sein. Letztendlich liegt es jedoch in der Zuständigkeit der AS-Betreiber ein Protokoll auszuwählen und dieses umzusetzen.

Sowohl die Separation als auch Elimination lösen die diskutierten Probleme in den Routing-Tabellen der DMZ [10]. Beide beiten Ansatzpunkte, die für Mobile-IP genutzt werden können [2],[14]. Es gibt aber eine Reihe von Gründen, die für eine Umsetzung des Separationsansatzes sprechen.

Um das Locator-Identifier-Problem zu lösen, speichern beide Ansätze einen Zustand. Beim Eliminationsansatz mit shim6 werden die erreichbaren Adressen der Kommunikationspartner beim Endgerät gespeichert. Diese Speicherung auf der Netzwerkschicht widerspricht dem Prinzip eines zustandsfreien Internets. Der dezentral gespeicherte Zustand erschwert die Suche nach Unregelmäßigkeiten und ist für Betrachter nicht transparent. Der Seperationsansatz speichert die Zuordnung von EIDs und RLOCs zentral im Mapping-System und bietet dem gesamten Internet dieselbe Sicht auf

den Status. Ein weiterer Nachteil der Elimination ist, dass ihre grundlegenden Techniken auf IPv6 basieren. Erst durch IPv6 kann eine Netzwerkschnittstelle per Entwurf mit mehrer Adressen versehen werden. Auch shim6 stellt eine Erweiterung für IPv6 dar. Der Sperationsansatz bietet hier eine generischere Lösung, insbesondere durch die Verwendung von LISP. Aufgrund des Tunneling-Mechanismus müssen keine Annahmen über die Netzwerkebene getroffen werden. Dies unterscheidet LISP auch von Six/One, welches die IPv6-Extension-Header nutzt. IPv6-basierende Lösungen haben vor allem den Nachteil, dass eine flächendeckende Nutzung von IPv6 abgewartet werden muss. Der Separationansatz erschafft eine Trennschicht zwischen der DFZ und den Edge-Netzwerken. Dies Trennschicht kann in Zukunft genutzt werden, um Änderungen auf der Netzwerkschicht unabhängig vom jeweils anderen Teil durchführen zu können [10]. Die Einführung der RLOCs bietet die Möglichkeit die Präfixe in der DFZ neu zu zuordnen, um durch bessere Aggregation eine noch kleinere Routing-Tabelle zu erhalten. Die Einführung von Eliminationsprotokollen wie shim6 erfodert die Mitarbeit vom mehr Netzwerkbetreibern als bei den Seperationsprotkollen nötig sind. Werden die Edge-Netzwerkbetreiber gezwungen auf die Vorteile ihrer providerunabhänige Adressen verzichten, könnten sie dies im gewissen Maße als Entmüdigung warhnehmen. Da auch der Haupteil der Kosten bei den Betreibern der Edge-Netzwerke ensteht, sind weitere poltischbedingte Verzögerungen zu erwarten. Die Entwicklungen der letzten Jahre machen aber eine zügige Überarbeitung des Routing-Systems nötig.

Es zu erwarten, dass spätestens mit der flächendeckenden Durchsetzung von IPv6 und der damit verbundenen höhren Anzahl an Präfixen sich das Problem weiter verschärfen wird. Dies könnte zusätzlichen Druck auf die Akteure ausüben, und zu einer schnellen Entscheidung führen.

Literatur

[1] S. Deering. The map & encap scheme for scalable ipv4 routing with portable site prefixes. Presentation, Xerox PARC, Palo Alto, CA, USA, 1996.

[2] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis. Internet Draft: Locator/ID separation protocol (LISP). RFC, IETF, Fremont, CA, USA, sepember 2009.

[3] Internet Engineering Task Force. online document tracker. avaible at: http://datatracker.ietf.org/, visited: 9th january 2010.

[4] Internet Engineering Task Force. online group charter. avaible at: http://www.ietf.org/dyn/wg/charter.html, visited: 9th january 2010.

[5] Internet Research Task Force. Routing research group website. avaible at: http://www.irtf.org/rrg, visited: 9th january 2010.

[6] V. Fuller, T. Li, J. Yu, and K. Varadhan. RFC 1519: Classless inter-domain routing (CIDR): an address assignment and aggregation strategy. RFC, IETF, Fremont, CA, USA, sep 1993.

[7] J. Hawkinson and T. Bates. RFC 1930: Guidelines for creation, selection, and registration of an Autonomous System (AS). Technical report, IETF, Fremont, CA, USA, march 1996.

[8] G. Huston. Analyzing the Internet's BGP routing table. *The Internet Protocol Journal*, 4(1), 2001.

[9] Luigi Iannone and Olivier Bonaventure. On the cost of caching locator/id mappings. In *CoNEXT '07: Proceedings of the 2007 ACM CoNEXT conference*, pages 1–12, New York, NY, USA, 2007. ACM.

[10] Dan Jen, Michael Meisel, He Yan, Dan Massey, Lan Wang, Beichuan Zhang, and Lixia Zhang. Towards a new internet routing architecture:arguments for separating edges from transit core. In *SIGCoMM '08: Proceedings of the 2008 ACM SIGCOMM Wororkshop Hotnets*, New York, NY, USA, 2008. ACM.

[11] R. Malhorta. *IP Routing*. O'Reilly Media, Sebastopol, CA, USA, 1st edition, 2002.

[12] Laurent Mathy and Luigi Iannone. Lisp-dht: towards a dht to map identifiers onto locators. In *CoNEXT '08: Proceedings of the 2008 ACM CoNEXT Conference*, pages 1–6, New York, NY, USA, 2008. ACM.

[13] Xiaoqiao Meng, Zhiguo Xu, Beichuan Zhang, Geoff Huston, Songwu Lu, and Lixia Zhang. IPv4 address allocation and the BGP routing table evolution. *SIGCOMM Comput. Commun. Rev.*, 35(1):71–80, 2005.

[14] E. Nordmark and M. Bagnulo. RFC 5533: Shim6: Level 3 multihoming shim protocol for IPv6. RFC, IETF, Fremont, CA, USA, june 2009.

[15] Andrew S. Tanenbaum. *Computer networks*. Prentice Hall, Upper Saddle river, NJ, USA, 4th edition, 2003.

[16] S. Thomson, T. Narten, and T. Jinmei. RFC 4862: IPv6 stateless address autoconfiguration. RFC, IETF, Fremont, CA, USA, september 2007.

[17] Christian Vogt. Six/one router: a scalable and backwards compatible solution for provider-independent addressing. In *MobiArch '08: Proceedings of the 3rd international workshop on Mobility in the evolving internet architecture*, pages 13–18, New York, NY, USA, 2008. ACM.