HLP: A Next-generation Interdomain Routing Protocol

Roman Golovatenko
(golovatenkoroman@mail.ru)
Technische Informatik M.Sc.

Seminar “Internet Routing”,

Technische Universität Berlin

Sommersemester 2009

Intelligent Networks / Intelligente Netze (INET)

Zusammenfassung

In dieser Arbeit wird ein Inter-Domain-Protokoll - “*Hybrid-Link-State-Path-Vektor-Protokoll*“ (HLP) eingeführt. Zuerst wird “*The Border Gateway Protocol*” (BGP) und dessen Schwierigkeiten analysiert. Dann wird HLP beschrieben, das bessere Eigenschaften wie Skalierbarkeit, Isolation und Konvergenz als BGP hat. Danach werden die Vorteile von HLP und anschließend die Lösung der Probleme von BGP bei HLP betrachtet. Zum Schluss wird eine Analyse der Eigenschaften von HLP und eine praktische Implementierung vorgestellt.

1.Einleitung

Das schnelle Wachstum des Internets stellt für Inter-Domain-Protokolle eine wichtige Herausforderung dar. Derartige Protokolle müssen wichtige algorithmische Eigenschaften wie Skalierbarkeit, Robustheit und schnelle Konvergenz erfüllen. Aus wirtschaftlichen Gründen sollen sie auch *Routing-Policys* unterstützten. Das führt zu einem Konflikt zwischen den wirtschaftlichen Interessen der privaten Routing-Policys und der strukturellen Notwendigkeit an robusten Routing-Algorithmen. In BGP sind fast alle Routing-Policys privat, deswegen leidet BGP unter den Problemen wie beispielsweise niedriger Skalierbarkeit, minimaler Fehlerisolation und langsamer Konvergenz. Die langsame Konvergenz entsteht aus der uninformierten Pfad-Exploration. In der Vergangenheit ist dies nicht so wichtig gewesen. Heute müssen diese Probleme wegen des enormen Internetwachstums gelöst werden. Um die Nachteile von BGP zu verringern, wurde ein neues Routingprotokoll - *Hybrid-Link-State-Path-Vektor-Protokoll* (HLP) entwickelt. Der Grundgedanke von HLP ist, dass Standard-Policys zu Verfügung gestellt werden, aber einige Pfad-Informationen vorenthalten werden. Die Standard-Policys basieren auf der Vermutung, dass die Internetrouten oft der Struktur der autonomen Systemhierarchie folgen wie sie durch die Provider-Kunden-Beziehungen geregelt ist. Die Kernidee, die für die Optimierung der Standard-Policys in HLP benutzt wird, ist, dass ausführliche Informationen von unnötigen Route-Updates über die Provider-Kunden-Hierarchie verborgen bleiben. Dazu wird die Sichtbarkeit der Routinginformation begrenzt. Die Unsichtbarkeit dieser Routinginformation in HLP führt zur wesentlichen Verbesserung der Skalierbarkeit, Isolation, Konvergenz und Fehlerdiagnose. HLP unterstützt nicht nur die meisten Policys von BGP, sondern führt auch eigene Policys ein. HLP ersetzt die Methode der Fragmentierung von Prefixen, die in BGP für das Traffic-Engineering benutzt wird, durch eine eigene Methode. Diese basiert auf dem Cost-Based-Traffic-Engineering und statischer Fragmentierung von Prefixen. Dabei werden die Probleme der Sicherheit und Fehlerdiagnose von BGP gelöst.

In Kapitel 2 dieser Arbeit werden die Probleme von BGP und allgemeine Designprobleme des Inter-Domain-Routingprotokolls beschrieben. Eine Beschreibung von HLP wird im Kapitel 3 geliefert. In Kapitel 4 und 5 werden eine Analyse der Eigenschaften und eine Implementierung von HLP dargestellt. Zum Schluss folgt ein Fazit.

2. Probleme von BGP

Das *Border Gateway Protocol* (BGP) ist ein Routingprotokoll des Internets. Es beschreibt, wie Router untereinander die Verfügbarkeit von Verbindungswegen zwischen autonomen Systemen (AS) weitergeben. BGP ist ein Pfadvektorprotokoll. Seine Funktionsweise ist stark an Distanzvektoralgorithmen und - protokollen wie z. B. Routing-Information-Protocol (RIP) angelehnt, jedoch wird dem dort vorkommenden Problem der Routingschleifen effektiv vorgebeugt. Ein BGP-Router teilt beim Senden von Verfügbarkeitsinformationen (Updates) dem Kommunikationspartner nicht nur mit, dass er einen bestimmten Abschnitt des Internets erreichen kann, sondern auch die komplette Liste aller ASe, welche die IP-Pakete bis zu diesem Abschnitt passieren müssen.

Die wichtigsten Eigenschaften für Inter-Domain-Routingprotokolle sind Skalierbarkeit, Isolation, Konvergenz und Routen-Stabilität. BGP hat Probleme mit diesen Eigenschaften:

- **Skalierbarkeit**. Das Inter-Domain-Routingprotokoll der Zukunft soll das enorme Wachstum des Internets berücksichtigen. BGP hat diesen Test nicht sehr gut bestanden, weil dessen Routing-State und Churn- Frequenz („*Churn*“ bedeutet: die Gesamtzahl von Updates, die durch eine Ereignis generiert werden) mit der Größe des Netzwerkes linear

wachsen; z.B. seit 1997 Jahr ist die Routing-Tabelle deutlich gewachsen: Die Anzahl der Einträge lag im Jahr 1997 bei ca. 50.000 Einträgen für über 3.000 AS, Ende 2005 bei ca. 170.000 Einträgen für über 26.000 Autonomen Systeme [1], und Ende Juni 2009 schon bei ca. 300.000 Einträgen für über 30.900 AS [6].

- **Isolation**. Kein Design von Inter-Domain-Routingprotokoll kann robust und skalierbar sein, wenn lokale Ereignisse global sichtbar sind. Die Analyse zeigt, dass BGP schlechte Eigenschaften der Fehlerisolation hat, weil ungefähr 20% des Routingereignisses global sichtbar sind und viele Routing-Updates werden durch Ereignisse, die weit vom Router entfernt sind, herbeigeführt[1].
- **Konvergenz und Routing-Stabilität**. Um eine gute Erreichbarkeit zu erreichen, sollen die Internetrouten relativ stabil sein und bei den notwendigen Änderungen einen neuen Zustand schnell annehmen. BGP leidet unter Route-Instabilitäten und langen Konvergenz-Zeiten.

Außerdem gibt es auch einige Designprobleme, die eine Modifizierung von BGP erfordern. Das sind Routingstruktur, Policys, Routing-Detallierungsstufe und Art des Routing-Protokolls.

- **Routingstruktur**. BGP zeigt die komplette Pfad-Information, weshalb lokale Routingereignisse global sichtbar gemacht werden. Das führt zur Verschlechterung der Skalierbarkeit und macht diese Ereignisse im Prinzip schwer isolierbar. Das bedeutet, dass ein Konfigurationsfehler den Rest des Netzwerkes beeinflussen kann.
- **Policys**. In BGP sind Policy-Informationen privat. Fast alle Beziehungen zwischen autonomen Systemen können als Peers, Kunden oder Providers kategorisiert werden, wobei diese Provider-Kunden-Beziehungen genau definiert werden können. Grundsätzlich läuft die Export-Regel jedoch immer nach folgenden Regeln ab, die auf diesen Inter-AS Beziehungen basieren: die Kunden-Routen müssen vor anderen Provider-Routen bevorzugt werden und die Routen von einem Peer oder Provider werden nicht einem anderen Peer oder Provider angeboten. Die Weigerung der Provider, diese Regeln offen zulegen, bedeutet, dass BGP zwischen einer falsch konfigurierten Policy und einer echten Policy nicht unterscheiden kann. Dadurch ist BGP kompliziert zu verwalten und zu diagnostizieren und ist empfindlich gegenüber Falschkonfigurationen und Angriffe.
- **Routing-Detallierungsstufe**. Die Tatsache, dass BGP ein prefixbasiertes Protokoll ist, verursacht ein Wachstum des Netzwerks eine enorme Zunahme der Prefixe in der Routing-Tabelle. Dieses Wachstum der Anzahl der Prefixe hat zu einer außerordentlichen Vergrößerung des Churn geführt.
- **Art des Routing-Protokolls**. BGP ist ein Pfadvektorprotokoll. Pfadvektor-Routing ermöglicht komplizierte Policys und leichte Schleife-Vermeidung. Der Nachteil ist, dass im Grenzfall die Konvergenz des Pfadvektorprotokolls mit der Länge des Pfads exponentiell wächst.

3. Das Routing-Modell von HLP

Der Entwurf von HLP basiert auf einer hierarchischen Struktur der AS-Topologie. Abbildung 1 illustriert diese einfache AS-Topologie mit verschiedenen Provider-Kunden-Hierarchien. Vereinfacht nimmt man an, dass jede Hierarchie nur auf die basischen Provider-Kunden-Beziehungen basiert und keine komplexen Beziehungen enthaltet. Ein autonomes System mit mehreren Providern kann ein Teil mehrerer Provider-Kunden- Hierarchien sein. Autonomen-Systeme verschiedener Hierarchien können miteinander auf verschiedenen Ebenen verbunden werden. Für die Verbindung werden Peering-Links benutzt.

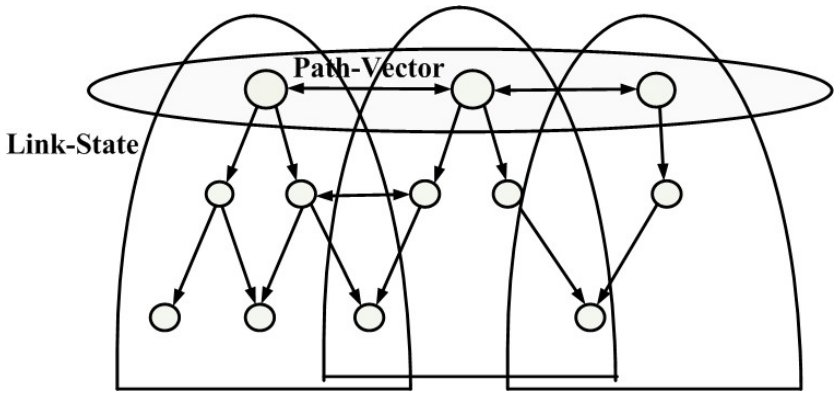
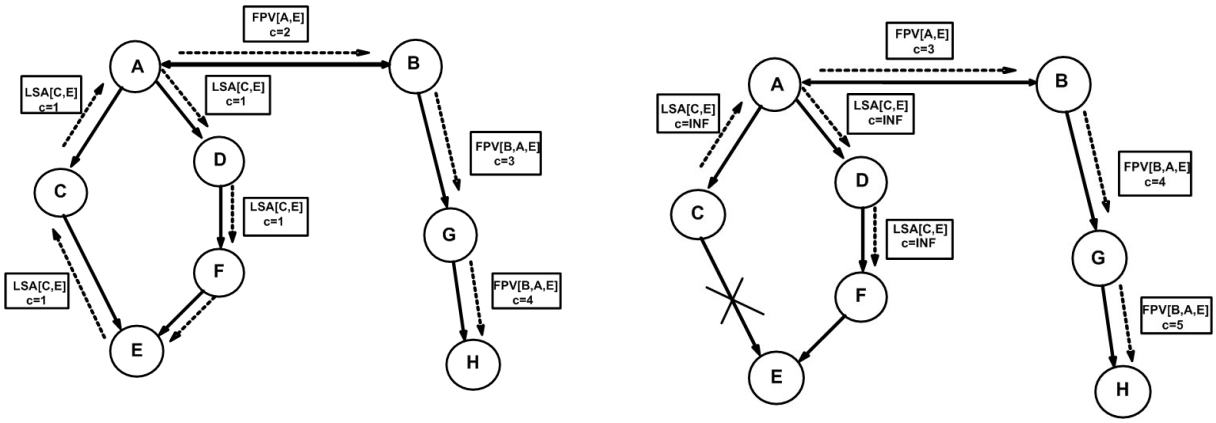


Abbildung 1: *AS-Topologie verschiedener Provider-Kunden- Hierarchien* (*Quelle* [1])

HLP verwendet eine Kombination aus Link-State-Routing und Pfadvektor-Routing. Link-State-Routing wird innerhalb einer Hierarchie benutzt. Jedes AS verteilt Link-State-Information über die Inter-AS Links innerhalb seiner eigenen Hierarchie und erneuert diese Information nach Empfang einer neuen Link-State-Nachricht. Pfadvektor-Routing wird zwischen Hierarchien benutzt und ein Pfadvektorteil von HLP ist gleich dem in BGP. Der primäre Unterschied ist, dass HLP einen geteilten Pfadvektor verwendet, der nur einen Teil des AS-Pfads zum Bestimmungsort enthält und einen Teil des AS-Pfads innerhalb AS-Hierarchie weglässt.

Betrachten wir in einem konkreten Beispiel, wie die Route von HLP innerhalb und zwischen AS-Hierarchien realisiert wird. Abbildung 2 zeigt zwei AS-Hierarchien: „A“ und „B“. Jeder Knoten unterstützt eine Datenbank mit Link-State-Informationen und eine Routing-Tabelle von Pfadvektoren. Knoten tauschen zwei Typen von Nachrichten aus: Link-State-Anzeigen (LSA) und Fragmented-Path-Vectors (FPV). Am Anfang werden alle Knoten von Hierarchie „A“ in LSA über die Existenz und die Kosten von Link (C, E) informiert. A bekommt eine LSA und sendet einen Pfadvektor nach B mit FPV (A, E) und Kosten 2. Dann wird der Pfadvektor ohne weitere Modifizierung des Pfads in Hierarchie „B“ nach dem Knoten H weiter geschickt. Abbildung 2-b illustriert den Fall, dass der Link (C, E) unterbrochen ist. Knoten der Hierarchie „A“ bekommen LSA, dass dieser Link nicht mehr existiert. Wenn A einen alternativen Weg kennt, dann schickt er nach B ein Pfadvektor-Update mit modifizierten Kosten. Dann schickt B FPV nach dem Knoten H. Das ist vergleichbar mit so genannten *Withdrawals* in BGP (*Withdrawals* bedeutet, dass Router sich mit Hilfe von Updates den Wegfall bestehender Routen mitteilen). Wenn A keinen alternativen Weg kennt, dann wird Route-Withdrawal zu B propagiert. FPV-Nachricht kann über mehr als ein Peering-Link verteilt werden. Diese Weiterleitung ermöglicht HLP indirektes Peering zu machen. In diesem Fall wird der FPV-Pfad alle Peering-AS enthalten, die in dessen Weg sind, oder die Kosten werden auf unendlich gesetzt.



(a) (b)

Abbildung 2: *Das Routing-Modell von HLP* (*Quelle* [1])

Die Zusammenfassung des HLP Modells ist:

- Alle ASe unterstützen eine Datenbank der Link-State-Topologie in ihrer lokalen Hierarchie.
- FPV fasst die Peering-Links zusammen, schließt aber die Teile des Pfads innerhalb der Hierarchien aus.
- Alle Inter-AS Links haben Kosten, die zu den Kosten des FPV hinzugefügt werden.

Die Hauptidee von HLP ist, dass nicht die ganze Information zwischen verschiedenen AS geteilt werden muss. Um die Informationsunsichtbarkeit zu erreichen, benutzt HLP das Konzept der Kostenverbergung. Im HLP werden drei Formen der Kostenverbergung unterstützt:

a) kleine Kostenänderungen der Kundenroute zwischen den Peering-Links sollen nicht geschickt werden;

b) kleine Kostenänderungen der Peerrouten sollen nicht zu den Kunden geschickt werden;

c) wenn zwischen zwei ASes, welche mehrere parallele Links untereinander haben, ein Link „unterbrochen“ ist, dann soll diese Information versteckt werden.

Die ersten zwei Fälle (a und b) werden in der Abbildung 3 gezeigt. In Abbildung 3-a können wir die erste Form der Kostenverbergung beobachten. Wenn Link (E, D) in der Hierarchie „A“ unterbrochen ist, dann wird ein alternativer Weg innerhalb dieser Hierarchie ausgewählt. Es werden keine Kostenänderungen der Kundenroute nach Hierarchie „X“ geschickt. Abbildung 3-b illustriert die zweite Forme der Kostenverbergung. In Hierarchie „X“ ist der Link zwischen den Knoten E und D unterbrochen. In diesem Fall wählt Knoten A den anderen Weg durch Hierarchie „Y“, um Knoten D zu erreichen, sendet aber keine Kostenänderungen der Peerrouten innerhalb der selben Hierarchie zu seinen Kunden.

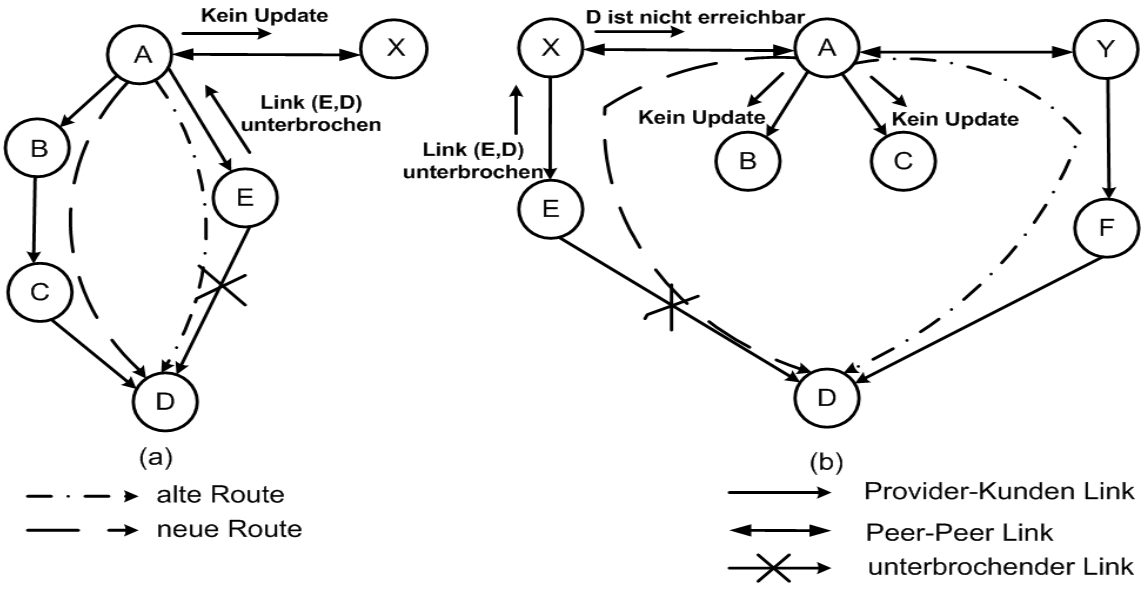


Abbildung 3: *Zwei Formen der Kostenverbergung* (*Quelle* [1])

Die Standard-Policys benutzen die Annahme, dass alle AS den nächststehenden Regeln folgen:

- Wege von einem Peer/Provider werden nicht an einen anderen Peer/Provider weitergegeben;
- Kundenwege müssen vor Providerwegen bevorzugt werden.

Wenn diese Regeln verletzt werden, dann wird eine Ausnahme in HLP ausgelöst. Es gibt zwei Formen von Ausnahmen:

1. Die Export-Policy: Wege von einem Peer/Provider können an einen anderen Peer/Provider weitergegeben werden.
2. Bevorzugung von Kunden-Routen: Man bevorzugt Nichtkundenwege vor Kundenwegen.

Messungen haben gezeigt, dass nur in 1% der Routen diese oben genannten Ausnahmen vorkommen.

4. Analyse von HLP

Für die Analyse der Eigenschaften von HLP wurde ein Route-Update-Emulator erstellt. Das Ziel dieses Emulators ist die Routing-Updates, welche durch ein einzelnes Ereignis ausgelöst werden, genau zu verfolgen. Die Eingabe für den Emulator sind eine AS-Topologie und eine Menge von Inter-AS Beziehungen. Es wird angenommen, dass es zwei Typen von Beziehungen gibt: “Provider-Kunde“ und “Peer-Peer“. Um Skalierbarkeit und Isolation zu vergleichen, wird die Analyse auf den Ausfall eines Inter-AS Links beschränkt. Die Analyse wurde auf einer echten AS-Topologie durchgeführt, die von RIPE [4] und Route-Views gesammelt wurde und 16774 AS und 37066 Inter-AS Links enthält. Dafür wurden zufällig 10 000 Inter-AS Links gewählt und anschließend unterbrochen. [1]

Die Analyse der Eigenschaften von HLP zeigt eine Verbesserung im Vergleich zu BGP. Man kann eine Verbesserung in der Skalierbarkeit, Isolation und Konvergenz erkennen.

An erste Stelle kann man die Verminderung der Churn- Frequenz von Route-Updates nennen. Diese Churn-Verminderung im HLP wurde durch zwei Faktoren erreicht. Die Faktoren sind: die Verwendung von Mapping, welche auf AS-Prefixen basiert und die Kostenverbergung von Route-Updates. Folgende Ergebnisse werden in ener Studie [1] gezeigt. Erstens, entsteht durchschnittlich 2% von Churn des BGP in HLP. Das illustriert die Verminderung der Churn- Frequenz um den Faktor 50. Zum zweiten, ist das entstehende Churn für 50% der Inter-AS Links um 75-mal geringer als im BGP [1]. Zum dritten, hängt die Churn-Verminderung von Typ der unterbrochenen Inter-AS Links ab. Die Abbildung 4 illustriert das Ausmaß des Churn in HLP und BGP. Wenn man den Kurvenverlauf betrachtet, fällt auf, dass der Churn in BGP bis 10 000 AS sich fast nicht verändert. In HLP zeigt der Churn ein deutliches Wachstum mit steigender AS-Anzahl.

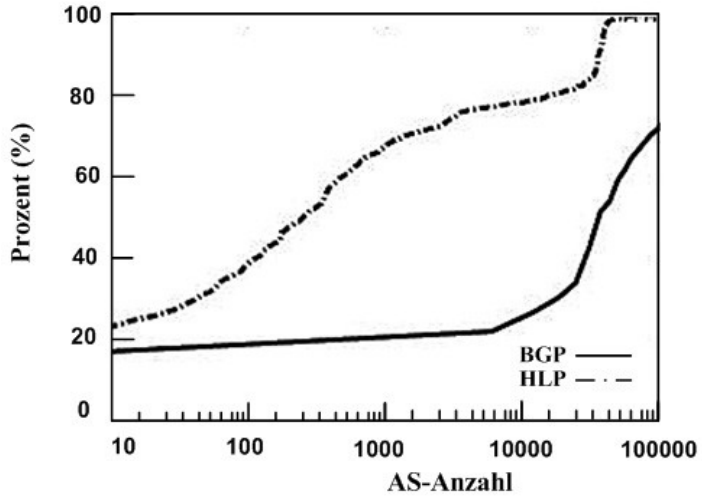


Abbildung 4: *Churn in HLP und BGP* (*Quelle* [1])

An zweiter Stelle soll die Verbesserung der Isolation genannt werden. Die Beobachtungen wurden gezeigt: HLP isoliert den Effekt der Routing-Ereignisse um 100-mal besser als BGP. Ungefähr 80% der Routing-Ereignisse im BGP sind global sichtbar. Im HLP wirken 40% der Ereignisse auf weniger als 10 AS [1]. Abbildung 5 zeigt das Ausmaß der Isolation, welches in HLP im Vergleich zu BGP erreicht wird. Wir können sehen, dass die Isolation in HLP mit Zunahme von ASen deutlich steigt. Bei BGP sehen wir, dass die Isolation bis 1000 AS sehr niedrig bleibt und erst ab 3000 AS stark ansteigt.

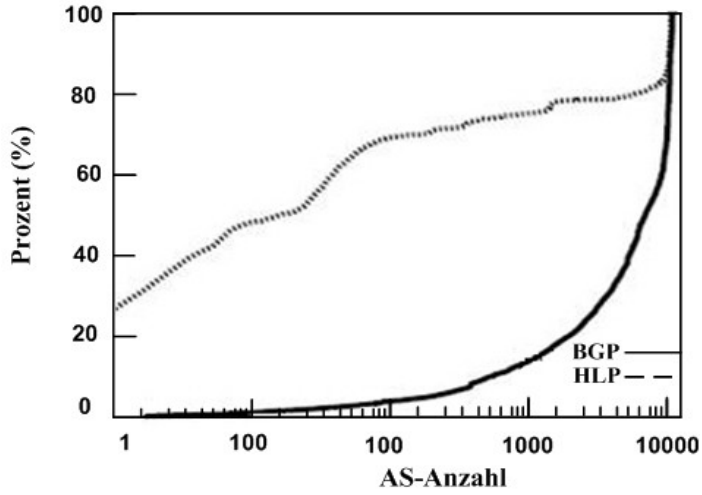


Abbildung 5: *Isolation in HLP und BGP* (*Quelle* [1])

Es gibt ein sehr interessantes Phänomen. Die Eigenschaften der Skalierbarkeit und Isolation verbessern sich mit dem zunehmenden Einsatz von Multihoming im Internet. Das passiert aufgrund der Existenz von alternativen Wegen bei Multihouming. Abbildungen 6a und 6b illustrieren eine Ereignisdistribution von Churn- und Isolation- Faktoren für verschiedene Arten von Inter-AS Links. Wir können sehen, dass der mittlere (50%) Faktor von Churn-Verminderung und der mittlere Isolation-Faktor für Multihomed- Kunden-Links ungefähr 200 und 1000 sind. Diese Faktoren für Tier1-Tier1 Links sind relativ klein, weil diese Links normalerweise in vielen Teilen geteilt werden und deshalb sehr schwer zu verbergen sind.

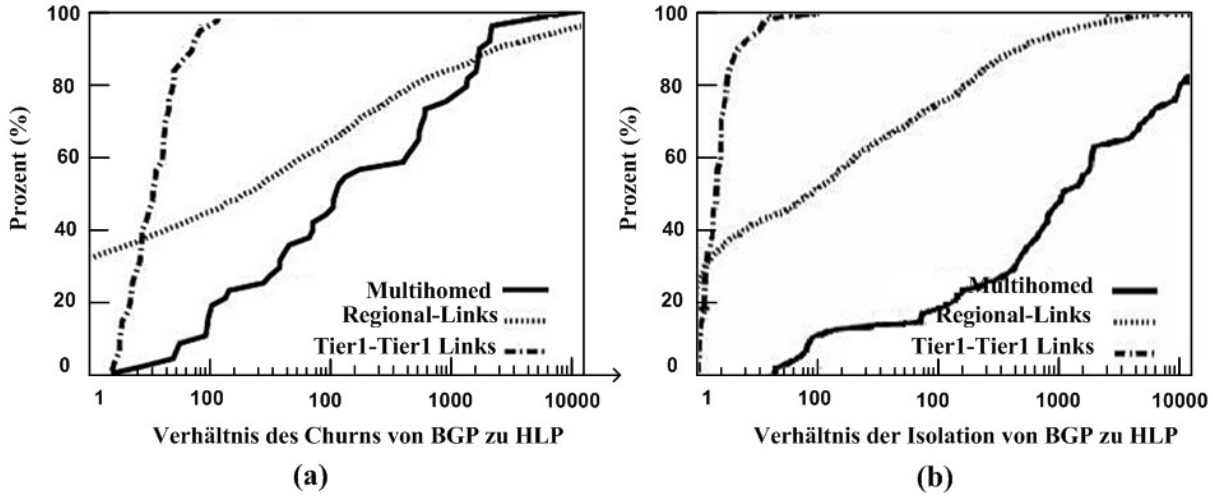


Abbildung 6: Ereignisdistribution *von Churn- und Isolation- Faktoren für verschiedene Arten von Inter-AS Links in HLP* (*Quelle* [1])

Die Konvergenz-Zeit ist der dritte Parameter, der durch HLP verbessert wird. Wie es in Quelle [3] gezeigt wird, verändert sich die Konvergenz-Zeit des Systems mit HLP mehr als linear im Vergleich zu BGP.

5. Die Praktische Implementierung von HLP

HLP wurde als ein Modul implementiert, das auf dem eXtensible Open Router Platform (XORP) als Software-Router basiert [5]. Die allgemeinen Operationen im HLP sind die Verarbeitung der LSA- und FPV-Updates. Die Tabellen 1 und 2 illustrieren die Ergebnisse der Implementierung. Tabelle 1 zeigt, wie die LSA-Bearbeitungszeit mit der Zunahme der Größe von AS-Hierarchie steigt; z.B. dass die LSA-Bearbeitungszeit für eine AS-Hierarchie mit der Größe 1000 0,052 Sek. ist. Tabelle 2 zeigt die Zunahme der FPV- Verarbeitungsrate mit steigender Anzahl von ASen. Wir können sehen, dass FPV-Verarbeitungsrate für eine AS-Hierarchie mit der Größe 20000 1132 Updates pro Sekunde ist.

Größe der AS-Hierarchie	100	300	500	700	1000
LSA-Bearbeitungszeit	0.0052	0.0153	0.0252	0.037	0.052

Tabelle 1: *Größe der AS-Hierarchie Vs. LSA Bearbeitungszeit (Sek.)* (*Quelle* [1])

Anzahl von ASen	1000	5000	10000	15000	20000
FPV-Verarbeitungsrate	5270	3154	1989	1452	1132

Tabelle 2: *Anzahl von ASen Vs.* FPV *Verarbeitungsrate (Updates/Sek.)* (*Quelle* [1])

Die Messungen wurden auf einem Rechner mit einem 2.4 GHz Intel Prozessor und 1 GB Speicher durchgeführt. Die Analyse dieser Messungen zeigt Folgendes:
Erstens, ist die Zahl von LSA innerhalb einer gegebenen Sekunde sehr klein im Vergleich zu BGP, welches viele Updates durch ein einziges Ereignis generiert. Zweitens ist die FPV-Verarbeitungsrate um 10-mal größer als es die heute maximale Update-Rate in BGP Routen ist.

6.Fazit

Das Wachstum des Internets in den letzten Jahren stellt an das Inter-Domain-Protokoll recht hohe Anforderungen. Die Frage ist, wie lange BGP sich an das Internetwachstum anpassen kann. Wie im Kapitel 4 dargestellt, zeigt HLP deutliche Verbesserung bei der Skalierbarkeit, Isolation, Konvergenz und Fehlerdiagnose im Vergleich zu BGP; beispielsweise weil HLP einige Pfad-Information versteckt, werden kleinen Änderungen im Routing-Verhalten in einer Hierarchie vor den Knoten anderer Hierarchien verborgen. Das führt zur Lösung des Problems der mangelnden Isolation von BGP. Andere praktische Beispiele von Vorteilen von HLP sind: HLP kann die Churn- Frequenz der Route-Updates um den Faktor 400 reduzieren und den Effekt von Routing-Ereignissen um 100-mal besser isolieren. Es zeigt, dass es sinnvoll ist HLP weiter zu entwickeln. Zurzeit ist HLP noch nicht bereit, um BGP zu ersetzen, aber dieses Protokoll hat unter der Voraussetzung weiterer Entwicklung gute Chancen eine Alternative für BGP zu werden.

Abkürzungsverzeichnis

AS - Autonome System

BGP - The Border Gateway Protocol

FPV - Fragmented-Path-Vector

HLP - Hybrid-Link-State-Path-Vektor-Protokoll

LSA - Link-State-Anzeige

XORP - eXtensible Open Router Platform

Literaturverzeichnis

[1] Lakshminarayanan Subramanian, Matthew Caesar, Cheng Tien Ee, Mark Handley, Morley Mao, Scott Shenker, Ion Stoica. *HLP: A Next-generation Interdomain Routing Protocol***.** Sigcomm 2005.

[2] Jennifer Rexford, Nick Feamster, Hari Balakrishnan. *Some Foundational Problems in Interdomain Routing.* Sigcomm 2005.

[3] Labovitz, C., Ahuja, A., Bose, A., and Jahanian, F. *Delayed Internet Routing Convergence.* In *Proc. ACM SIGCOMM* (2000).

[4] RIPE's Routing Information Service Raw Data Page, Website, http://data.ris.ripe.net/

[5] The eXtensible Open Router Platform (XORP), Website, http://www.xorp.org

[6] Potaroo, Website, http://www.potaroo.net/tools/asn32/