

# Bestimmung von Daten Pfaden auf BGP AS Ebene

Harald Schiöberg

17. Januar 2005

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>
<b>2</b>	<b>Werkzeuge</b>
2.1	Reverse DNS . . . . .
2.2	traceroute . . . . .
2.3	whois . . . . .
2.4	Looking Glass Server . . . . .
<b>3</b>	<b>Methoden</b>
3.1	AS Pfad . . . . .
3.2	Internet Route Registry . . . . .
3.3	Origin AS . . . . .
3.4	Erweitertes originating AS . . . . .
3.5	Vergleich der Methoden . . . . .
<b>4</b>	<b>Auflösung unvollständiger Pfade</b>
4.1	Nicht aufgelöste Hops in einem AS . .
4.2	Nicht zugeordnete Hops zwischen mehreren AS . . . . .
4.3	Multi Origin AS . . . . .
<b>5</b>	<b>Weitere Verbesserungen des IP-AS Mappings</b>
5.1	Typische Muster von Abweichungen . .
5.2	Internet Exchange Points (IXPs) . . . .
5.3	Verwandte AS . . . . .
5.4	Adressen ohne BGP Eintrag . . . . .
5.5	Ergebnis . . . . .
<b>6</b>	<b>Tatsächlich nicht übereinstimmende Pfade</b>
6.1	Filter und Aggregate . . . . .
6.2	IP Adressen an AS Grenzen . . . . .
6.3	ICMP Quelladressen . . . . .
6.4	Routing Anomalien . . . . .
<b>7</b>	<b>Ergebnisse</b>

## 1 Einleitung

1 Traceroute ist ein mächtiges Werkzeug zur Fehlersuche bei Routinganomalien im Internet. Der nächste Schritt nach der Erkennung sollte jedoch das Auffinden geeigneter Ansprechpartner sein, um die Störung möglichst schnell zu beheben. Dies stellt sich als eine erstaunlich komplexe Aufgabe dar. Ein Werkzeug, das diese Zuordnung automatisiert, könnte im Fehlerfall viel kostbare Zeit und sparen.

2 Des weiteren ist ein Werkzeug, das die Übereinstimmung von Routing Daten (in diese Fall der BGP Tabelle) mit dem tatsächlichen Verkehrs Fluss validiert ein wertvolles Hilfsmittel für den Administrator von Netzwerk Komponenten im Internet.

4 In der Forschung bauen viele Arbeiten auf der Kenntnis der Internettopologie auf. Häufig wird dabei die Topologieinformation aus der BGP Tabelle gewonnen. Wie sich zeigen wird, ist diese Topologieinformation nicht immer korrekt, auch hier würde eine sichere Validierung die Grundlage vieler Arbeiten sichern.

5 In dieser Arbeit soll die Entwicklung eines solchen AS Level Traceroute Tools vorgestellt werden. Dazu werden im Abschnitt 2 einige grundlegende Werkzeuge vorgestellt, auf deren Kenntnis das weitere Verständnis aufbaut. In Abschnitt 3 werden die Methoden zur Pfad-Erkennung beschrieben. In Abschnitt 4 werden noch einige Methoden zur Behandlung von Spezialfällen vorgestellt. Schließlich werden mit Abschnitt 6 die Ursachen untersucht, warum auch dieses Werkzeug keine perfekten Ergebnisse liefern kann. (FIXME eine Section fehlt)

8 Diese Arbeit basiert auf dem Artikel „Towards an Accurate AS-Level Traceroute Tool“ von Zhuoqing Morley Mao, Jennifer Rexrod, Jia Wang und Randy H. Katz; Proceedings of Sigcomm 2003. Sie vereinfacht viele Zusammenhänge, insbesondere sind die genauen Messergebnisse ausgespart.

## 2 Werkzeuge

Einige weit verbreitete Werkzeuge, die bei den hier vorgestellten Verfahren Verwendung finden, sollen hier kurz vorgestellt werden.

### 2.1 Reverse DNS

Die Hauptfunktion des Domain Name Service (DNS) ist es, Namen IP-Adressen zuzuordnen. Dies dient hauptsächlich der Vereinfachung der Handhabung, da aussagekräftige Namen (www.net.in.tum.de) leichter zu merken sind, als numerische IP Adressen (131.159.15.242). Es ist möglich, auch die umgekehrte Abfrage zu machen, also einer IP Adresse einen Namen zuzuordnen. Dies kann helfen, eine IP Adresse einem Provider zuzuordnen, da entweder der Name selbst eine Zugehörigkeit zu einem Provider ergibt, oder bei der zuständigen Name Registry erfragt werden kann, wem die entsprechende Domain zugeordnet ist.

### 2.2 traceroute

traceroute ist ein Werkzeug zur Bestimmung des forwarding Pfad im Internet. Dazu werden UDP Pakete mit inkrementell ansteigender „time to live“ (TTL) im IP Header an den Ziel-Rechner versandt. Da die TTL dieser Pakete an jedem Router decremementiert wird, wird sie vor Erreichen des Ziels zu Null. Der entsprechende Router muss nach RFC 791 das Paket verwerfen und mit einem Internet Control Message Protocol (ICMP) Paket, Typ 11, Code 0 (time to live exceeded in transit) (RFC 792) an den Absender antworten. Die Source Adresse dieses ICMP Pakets gibt die IP Adresse des forwarding Hops. Der traceroute Lauf kann beendet werden, wenn ein ICMP Paket Typ 3, Code 3 (Port Unreachable) vom Zielrechner empfangen wird.

Die meisten Router verhalten sich an dieser Stelle RFC konform. Sobald man jedoch die Provider-Netze verlässt, und in die Nähe von Endsystemen kommt, werden die Test-Pakete, und oft auch die ICMP Antworten, häufig von Firewalls blockiert. Die möglichen Antworten reichen dann von, „Paket wird kommentarlos fallen gelassen“ (der `bsd traceroute` zeigt das als \* an) bis „ICMP port unreachable“ von der Firewall.

### 2.3 whois

Ähnlich wie DNS stellt whois weitere Informationen

über IP Adressen und Domain Namen zur Verfügung. Mittels dieses Werkzeugs kann z.B. erfragt werden, wer für ein bestimmtes Netz verantwortlich, welchem AS es zugeordnet oder auf wen eine bestimmte Domain registriert ist. Dazu stellen die Internet Registries Server zur Verfügung, auf denen diese Information abgerufen werden kann. Die Provider sind angehalten, diese Daten stets auf dem aktuellen Stand zu halten. In der Praxis sind die whois-Einträge jedoch nur selten wirklich aktuell.

### 2.4 Looking Glass Server

Looking Glass Server sind öffentliche Server, die BGP Tabellen zur Verfügung stellen. Einige sind einfach router, auf die man sich per telnet einloggen kann, andere sind www Server, von denen man per Webfrontend die Routinginformation abrufen kann. Manche der web-basierten Looking Glass Server halten Routing-Information bereit, die an verschiedenen Beobachtungspunkten gewonnen wurde.

## 3 Methoden

### 3.1 AS Pfad

Eine nahe liegende Methode ist es, aktuelle BGP-Daten zu benutzen, um den AS-Pfad eines Paketes zwischen 2 Hosts zu bestimmen. Diese Methode kommt sogar ganz ohne Probe-Pakete aus, da der Pfad aus den Daten berechnet werden kann.

FIXME(Absatz genauer) Daraus ergibt sich das größte Problem: man benötigt aktuelle BGP-Daten. Diese sind schon theoretisch nahezu unmöglich zu beschaffen, man bedenke die großen Verzögerungen, die im BGP Protokoll auftreten. Wenn man zum Beispiel diese Daten aus dem nächsten eigenen Router ausliest, so hat man keinerlei Gewähr, dass sich auf dem Pfad nicht schon längst große Änderungen ergeben haben. Die einzige Möglichkeit, diese Daten wirklich aktuell zu bekommen, ist es die Routing-Tabellen aller Router auf dem Pfad zu einem genau definierten Zeitpunkt auszulesen. Eine Möglichkeit, an möglichst viele BGP-Daten zu kommen, sind öffentliche Looking Glass Server.

Außerdem hat diese Methode das Problem, dass sie BGP-Routen und Paket-Routen gleichsetzt. Das ist jedoch in der Praxis nicht gegeben, insbesondere sind derartige Werkzeuge ja gerade dann von Interesse,

Abbildung 1: Ausgabe von prtraceroute

```
prtraceroute to www.mit.edu (18.7.22.83), 30 hops max, 12 byte packets
1 [AS680] 131.159.14.126 (131.159.14.126) 1.158 ms 1.206 ms 0.948 ms
2 [AS680] nz-hv19sl-net.informatik.tu-muenchen.de (131.159.252.149) 2.110
3 [AS680] nz-bbl-hv19sl.informatik.tu-muenchen.de (131.159.252.46) 1.576
4 [AS680] nz-gate-bbl.informatik.tu-muenchen.de (131.159.252.6) 1.266 ms
5 [AS680] wangate.informatik.tu-muenchen.de (131.159.252.2) 1.481 ms 1.
6 [AS680] carwan.lrz-muenchen.de (129.187.1.2) 2.475 ms 2.058 ms 1.718
7 [AS680] ar-muenchen1-ge0-1-222.g-w.in.dfn.de (188.1.37.13) 1.774 ms 1.
8 [AS680] cr-muenchen1-ge0-0.g-w.in.dfn.de (188.1.74.1) 2.022 ms 1.992 m
9 [AS680] cr-frankfurt1-p0-3.g-w.in.dfn.de (188.1.18.26) 8.085 ms 8.084
10 [AS20965] dfn.dei.de.geant.net (62.40.105.1) 8.488 ms 8.367 ms 8.248
11 [AS20965] dei-2.de2.de.geant.net (62.40.96.53) 8.334 ms 8.429 ms 8.1
12 [AS20965] abilene-gw.de2.de.geant.net (62.40.103.254) 112.825 ms 102.
13 [AS0] nycmg-washng.abilene.ucaid.edu (198.32.8.84) 98.603 ms 98.646
14 [AS0] noxgsl-PO-6-0-NOX-NOX.nox.org (192.5.89.9) 103.679 ms 103.573 m
15 [AS0] noxgsl-PEER-NOX-MIT-192-5-89-90.nox.org (192.5.89.90) 103.85 ms
16 [AS3] B24-RTR-3-BACKBONE.MIT.EDU (18.168.0.26) 104.081 ms 104.007 ms
17 [AS3] WWW.MIT.EDU (18.7.22.83) 215.706 ms 104.114 ms 103.783 ms

Path taken:
AS680 AS20965 (???) (???) (???) AS3
17 AS3 WWW.MIT.EDU destination -> internal
16 AS3 B24-RTR-3-BACKBONE.MIT.EDU internal -> peer unregistered
15 AS0 noxgsl-PEER-NOX-NOX-MIT-192-5-89-90.nox.org registered -> registered
14 AS0 noxgsl-PO-6-0-NOX-NOX.nox.org registered -> registered
13 AS0 nycmg-washng.abilene.ucaid.edu registered -> registered
12 AS20965 abilene-gw.de2.de.geant.net import: 2 -> internal
11 AS20965 dei-2.de2.de.geant.net internal -> internal
10 AS20965 dfn.dei.de.geant.net internal -> export
9 AS680 cr-frankfurt1-p0-3.g-w.in.dfn.de import: 2 -> internal
8 AS680 cr-muenchen1-ge0-0.g-w.in.dfn.de internal -> internal
7 AS680 ar-muenchen1-ge0-1-222.g-w.in.dfn.de internal -> internal
6 AS680 carwan.lrz-muenchen.de internal -> internal
5 AS680 wangate.informatik.tu-muenchen.de internal -> internal
4 AS680 nz-gate-bbl.informatik.tu-muenchen.de internal -> internal
3 AS680 nz-bbl-hv19sl.informatik.tu-muenchen.de internal -> internal
2 AS680 nz-hv19sl-net.informatik.tu-muenchen.de internal -> internal
1 AS680 131.159.14.126 internal -> internal
0 AS680 penguin.net.informatik.tu-muenchen.de internal -> source
```

wenn es zu Anomalien kommt. Man kann zum Beispiel den Fall konstruieren, in dem BGP zwar auf eine stabile Route konvergiert ist, die Pakete aber aufgrund von anderen Routing-Entscheidungen (z.B. eines Interior-Protokolls) trotzdem nicht diesem Pfad folgen.

Man wird reine BGP-Daten kaum benutzen um den tatsächlichen Verkehrsfluss nachzuvollziehen. Allerdings kann die Korrelation von BGP-Routen mit tatsächlichen Routen gute Hinweise auf Anomalien im Internet-Routing liefern.

### 3.2 Internet Route Registry

Eine andere Methode ist es, die Internet Route Registries zu benutzen. Zu jedem Netzwerkprefix, welches im Internet geroutet wird, existiert bei der zuständigen Registry ein Eintrag, wem das Prefix „gehört“. Dort ist auch das jeweilige AS vermerkt. Diese Daten können von den öffentlichen whois Servern abgerufen werden. Daher kann man leicht ein traceroute Programm schreiben, das für jeden Hop ein whois query macht und die hops auf diese Weise dem jeweiligen AS zuordnet. Abbildung 3.2.

Der grosse Vorteil ist, dass man keine BGP Daten benötigt. So kann man von jedem Rechner aus sofort einen solchen traceroute Lauf machen. Die benötigten whois Daten stehen auf öffentlichen Servern zur Verfügung.

Das Hauptproblem dieser Methode ist die Aktualität der whois Datenbank. Da es sich dabei nicht um

Abbildung 2: Whois Abfrage für 131.159.74.65

```
OrgName: RIPE Network Coordination Centre
OrgID: RIPE
Address: P.O. Box 10096
City: Amsterdam
StateProv:
PostalCode: 1001EB
Country: NL

ReferralServer: whois://whois.ripe.net:43

NetRange: 131.159.0.0 - 131.160.255.255
CIDR: 131.159.0.0/16, 131.160.0.0/16
NetName: RIPE-ERX-131-159-0-0
NetHandle: NET-131-159-0-0-1
Parent: NET-131-0-0-0-0
NetType: Early Registrations, Transferred to RIPE NCC
Comment: These addresses have been further assigned to users in
the RIPE NCC region. Contact information can be found in
the RIPE database at http://www.ripe.net/whois
Regdate: 2004-02-04
Updated: 2004-02-04

# ARIN WHOIS database, last updated 2005-01-11 19:10
# Enter ? for additional hints on searching ARIN's WHOIS database.

Found a referral to whois.ripe.net:43.
% This is the RIPE Whois query server #2.
% The objects are in RPSL format.
%
% Rights restricted by copyright.
% See http://www.ripe.net/db/copyright.html

inetnum: 131.159.0.0 - 131.159.255.255
remarks:
remarks: This inetnum has been transferred as part of the ERX.
remarks: It was present in both the ARIN and RIPE databases, so
remarks: the information from both databases has been merged.
remarks: If you are the mntnr of this object, please update it
remarks: to reflect the correct information.
remarks:
remarks: Please see the information for this process:
remarks: http://www.ripe.net/db/erx/erx-ip/network-131.html
remarks:
remarks: **** INFORMATION FROM ARIN OBJECT ****
remarks:
remarks: netname: TUM-INFO-LAN
remarks: descr: Institut fuer Informatik der TU Muenchen
remarks: Arctisstrasse 21
remarks: Postfach 202420
remarks: 8000 Muenchen 2
remarks: country: DE
remarks: admin-c: HR838-RIPE
remarks: tech-c: HR838-RIPE
remarks: changed: hostmaster@arin.net 19881207
remarks: changed: hostmaster@arin.net 19970326
remarks: **** INFORMATION FROM RIPE OBJECT ****
remarks:
remarks: netname: TUM-INFO-LAN
remarks: descr: Institut fuer Informatik der TU Muenchen
remarks: Arctisstrasse 21; Postfach 202420. 80333 Muenchen
remarks: DE
remarks: admin-c: HR55
remarks: tech-c: HR55
remarks: EM11-RIPE
remarks: status: ASSIGNED PI
remarks: mnt-by: LRZ-MNT
remarks: mnt-routes: LRZ-MNT
remarks: changed: portens@gmd.de 19910820
remarks: changed: rv@informatik.uni-dortmund.de 19930203
remarks: changed: poldi@dfn.de 20021024
remarks: changed: poldi@dfn.de 20021122
remarks: changed: er-transfer@ripe.net 20040204
remarks: source: RIPE

route: 131.159.0.0/16
remarks: descr: TUM-INFO-LAN
remarks: origin: AS1216
remarks: mnt-routes: DFN-MNT
remarks: mnt-by: LRZ-MNT
remarks: changed: tfoeb@lrz.de 20021125
remarks: source: RIPE

route: 131.159.0.0/16
remarks: descr: TUM-INFO-LAN
remarks: origin: AS680
remarks: mnt-routes: LRZ-MNT
remarks: mnt-by: DFN-MNT
remarks: changed: helliger@noc.dfn.de 20000526
remarks: changed: poldi@dfn.de 20021125
remarks: source: RIPE

route: 131.159.0.0/16
remarks: descr: TUM-INFO-LAN
remarks: origin: AS1275
remarks: mnt-routes: LRZ-MNT
remarks: mnt-by: DFN-MNT
remarks: changed: ripe-dm@ripe.net 19941121
remarks: changed: poldi@dfn.de 19950907
remarks: changed: poldi@dfn.de 20021125
remarks: source: RIPE

person: Hans-Otto Riethmayer
address: Technische Universitaet Muenchen
address: Fakultaet fuer Informatik
address: Arctisstrasse 16
address: D-W-8000 Muenchen 2
address: Germany
address: +49 89 2105 2025
address: +49 89 2105 8232
e-mail: rietmayer@informatik.tu-muenchen.de
nic-hdl: HR55
changed: poldi@dfn.de 19930107
changed: rv@informatik.uni-dortmund.de 19930112
source: RIPE

person: Hans-Otto Riethmayer
address: Institut fuer Informatik der TU Muenchen
address: Arctisstrasse 21
address: Postfach 202420
address: 8000 Muenchen 2
address: US
address: +49 89 2105 2025
e-mail: rietmayer@informatik.tu-muenchen.de
nic-hdl: HR838-RIPE
mnt-by: RIPE-ERX-MNT
changed: hostmaster@arin.net 19970326
changed: er-transfer@ripe.net 20040105
source: RIPE

person: Elmar Bartel
address: Technische Universitaet Muenchen
address: Fakultaet fuer Informatik
address: Arctisstrasse 16
address: D-W-8000 Muenchen 2
address: Germany
address: +49 89 2105 2025
address: +49 89 2105 8232
e-mail: bartel@informatik.tu-muenchen.de
nic-hdl: EM11-RIPE
mnt-by: DFN-NTFY
changed: noc@noc 19950704
source: RIPE
```

für den eigentlichen Betrieb wichtige Daten handelt, ist die Motivation für die Netzbetreiber diese Daten aktuell zu halten sehr gering. Aus diesem Grund liefern diese Anfragen oft falsche oder gar keine Ergebnisse. In Abbildung 2 kann man das beobachten: Die Postleitzahlen sind 4-stellig, keine der Telefonnummern stimmt, und die Informatik ist immer noch in der Innenstadt.

### 3.3 Origin AS

Das originating AS ist dasjenige AS, das in einem BGP Pfad als letztes steht. Man kann nun das originating AS für jeden einzelnen traceroute Hop bestimmen, bzw. für das längste Prefix aus der BGP Tabelle, das die IP-Adresse des Hops enthält. Der größte Unterschied zur Berechnung aus BGP Pfaddaten besteht darin, dass man den forwarding Pfad mittels traceroute bestimmt und dann für jeden gemeldeten Hop auf die BGP Daten zurückgreift. Das heißt, man ist nun völlig unabhängig von den Fluktuationen im BGP Routing. Das originating AS sollte sich eigentlich nur ändern, wenn ein Netz umzieht, aber niemals durch Dynamiken des BGP Protokolls.

Da diese Daten Orts unabhängig sind, kann zu ihrer Gewinnung auch auf öffentliche Looking Glass Server zugegriffen werden (z.B.: <http://www.ris.ripe.net/cgi-bin/lg/index.cgi>).

Daten, die aus dem BGP Protokoll gewonnen werden, können nicht der Nachlässigkeit der Netzverantwortlichen unterliegen. Schließlich würde das Netz mit falschen Daten nicht funktionieren.

Ein Problem bei der Methode stellen so genannte Multi Origin AS dar, also solche Netze, die von mehreren AS announced werden. Diese sind nicht so selten, da z.B. die Adressen von Routern an Internet Exchange Points (IXP), die von verschiedenen Providern genutzt werden, oft von mehreren auch announced werden.

Ein anderes Problem stellen Netze dar, die nicht in der BGP Tabelle vorkommen. Dies kann bei Infrastrukturnetzen der Fall sein, und gerade diese sind ja besonders interessant. Eine besonders heimtückische Falle stellen Netze dar, die von ihrem Betreiber nicht announced werden, aber auf Grund eines größeren Aggregates trotzdem in der BGP Tabelle aufzutauchen scheinen.

### 3.4 Erweitertes originating AS

Für dieses Verfahren benötigt man Messpunkte im Internet, von denen man sowohl die BGP-Tabelle kennt, und die einen traceroute durchführen können. Man arbeitet die gesamte Routing-Tabelle ab. Aus jedem Prefix werden einige IPs ausgewählt, und ein traceroute auf diese Hosts durchgeführt. Der BGP Pfad wird auf Anomalien überprüft, und gegebenenfalls als unbrauchbar verworfen:

- *Kein AS Pfad* Bei Adressen, die im selben AS liegen wie der Messpunkt, gibt es keinen AS Pfad.
- *Private AS Nummern* AS Pfade können private AS Nummern enthalten. Dies deutet auf einen Fehler in der BGP Konfiguration hin.
- *Scheinbare Schleifen* BGP hat zwar einen Schleifenerkennungsmechanismus, aber scheinbare Schleifen können auftreten, wenn Router willkürliche AS Nummern dem Pfad voranstellen.
- *AS-SET* Wenn Router Prefixes aggregieren, geben sie die Menge aller AS Nummern der Teil-Prefixes als ungeordnete Menge im Pfad an. Hier lassen sich keine Aussagen über den wirklichen Pfad treffen.

Nun wird der traceroute AS Pfad berechnet, in dem jedem traceroute Hop ein AS zugeordnet wird. Das wird über das Originating AS der jeweiligen Hop-IP gemacht. Aus diesen, von mehreren Messpunkten gesammelten Daten wird eine IP-AS Tabelle erstellt, die später für die Zuordnung nach traceroute Läufen genutzt werden kann.

### 3.5 Vergleich der Methoden

Die größten Probleme der einzelnen Methoden sind in Tabelle 1 noch einmal zusammengefasst.

Die Ergebnisse lassen sich ungefähr so zusammenfassen: Vergleicht man BGP Pfad Informationen mit traceroute Daten, denen über die whois-Server ein AS zugeordnet ist, so erhält man in rund 45% der Fälle eine Übereinstimmung des AS Pfades. Die Übereinstimmung zwischen BGP Pfad und forwarding Pfad bei Gewinnung der IP-AS Zuordnung aus originating AS Informationen von mehreren Messpunkten liegt bei rund 70%.

Im weiteren soll nun besprochen werden, die verbleibenden Fehler weitgehend eliminiert werden können.

Tabelle 1: Probleme der einzelnen Methoden

	AS Pfad	Registries	Erw. Orig. AS
Aufwand	hoch	niedrig	sehr hoch
Aktiv	nein	ja	ja
BGP Daten	kritisch	keine	teilweise
whois Daten	n/a	kritisch	gering

## 4 Auflösung unvollständiger Pfade

### 4.1 Nicht aufgelöste Hops in einem AS

Antwortet ein Host nicht auf die traceroute Pakete, und werden vorheriger und nachfolgender Hop dem selben AS zugeordnet, so werden alle 3 Hosts als ein AS Hop angesehen.

### 4.2 Nicht zugeordnete Hops zwischen mehreren AS

Kann für einen traceroute Hop kein AS Mapping gefunden werden und liegt dieser Hop zwischen zwei AS, so wird versucht, diesen Hop mittels whois und reverse DNS Information einem Provider zuzuordnen.

### 4.3 Multi Origin AS

Gehört eine IP eines traceroute Hops zu einem Prefix, das von mehreren AS announced wird, so gilt jedes dieser AS als Match im Vergleich mit dem BGP Pfad.

## 5 Weitere Verbesserungen des IP-AS Mappings

Durch die bisher aufgeführten Methoden kann die Anzahl der Routen, in denen die BGP Route nicht mit der traceroute Route übereinstimmt, auf 6-9% reduziert werden. Die meisten dieser Fälle sind auf Ungenauigkeiten im IP-AS Mapping zurückzuführen.

### 5.1 Typische Muster von Abweichungen

Über zwei Drittel der Differenzen fällt in eine dieser vier Kategorien:

**Extra AS Hop** (Abbildung 3) 30-40% der Pfade mit Abweichungen haben einen zusätzlichen Hop im forwarding Pfad.

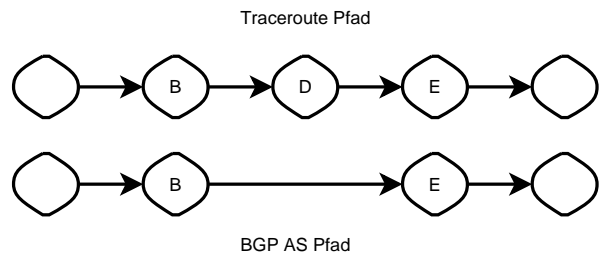


Abbildung 3: Extra AS Hop

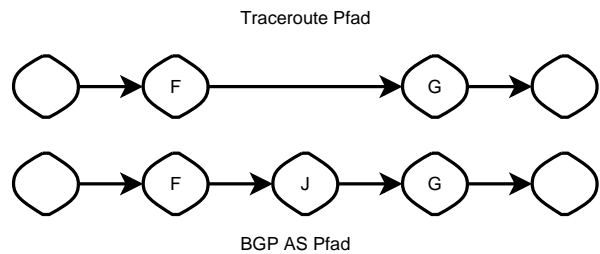


Abbildung 4: Fehlender AS Hop

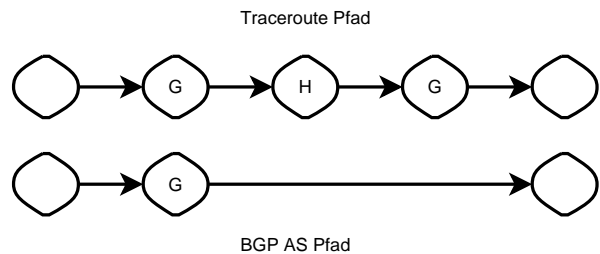


Abbildung 5: AS Schleife (Abstand 2)

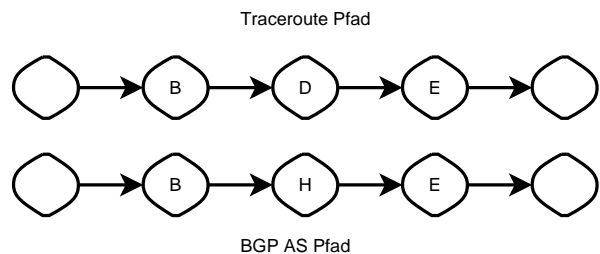


Abbildung 6: Abweichender AS Hop

**Fehlender AS Hop** (Abbildung 4) 20% der Abweichungen rühren von AS Pfaden, in denen ein AS im BGP zu fehlen scheint.

**AS Loop** (Abbildung 5) Rund 10% der traceroute AS Pfade haben eine Schleife auf dem AS-Level, bei dem ein AS zweimal im Pfad auftaucht, mit nur einem AS Hop dazwischen.

**Abweichender Hop** (Abbildung 6) In 2-3% der Fälle hatten die 2 Pfade ein unterschiedliches AS in der Mitte des Pfades.

Folgende Erklärungen gibt es für diese häufig auftretenden Anomalien.

## 5.2 Internet Exchange Points (IXPs)

Internet Exchange Points (IXPs) sind Austauschpunkte zwischen den Providern, an denen Daten von einem Provider zu einem anderen übergeben werden. An diesen Punkten befindet sich typischerweise ein leistungsfähiges lokales Netzwerk, an das die Router der einzelnen Provider angeschlossen sind. 2 Provider, die dort Daten austauschen wollen, etablieren über dieses lokale Netz hinweg eine direkte BGP Session.

Im BGP Pfad tauchen daher diese beiden Router als direkt benachbart auf, während im traceroute-Pfad die Adressen der lokalen Router als zusätzliche Hops sichtbar werden (vgl. Abbildungen 7 und 8). IXPs sind daher eine typische Ursache für zusätzliche AS Hops im forwarding Pfad (Abbildung 3). Diese Adressen können

- einem der angeschlossenen Provider
- mehreren der angeschlossenen Provider (MOAS)
- keinem der angeschlossenen Provider, sondern dem Betreiber des IXP

zugeordnet sein.

Es wird generell angenommen, dass im BGP Pfad kein IXP auftaucht. Im forwarding Pfad werden hingegen Hops auftreten, die einem IXP zugeordnet sind.

IXP kann man daran erkennen, dass viele forwarding Pfade unterschiedliche AS als Nachbarhops haben, da ja viele AS an einen IXP angeschlossen sind. Wenn man also IXPs auf diese Weise identifiziert, kann man einfach eine Liste der IXPs pflegen, und diese forwarding Hops beim Vergleich mit BGP Pfaden ignorieren.

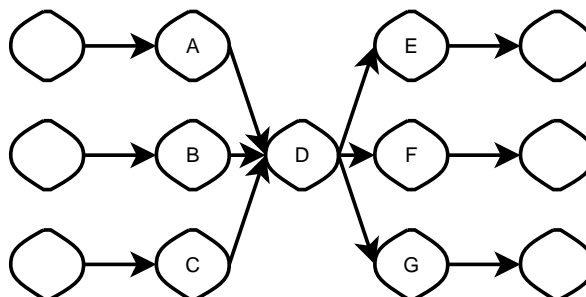


Abbildung 7: Forwarding AS Path

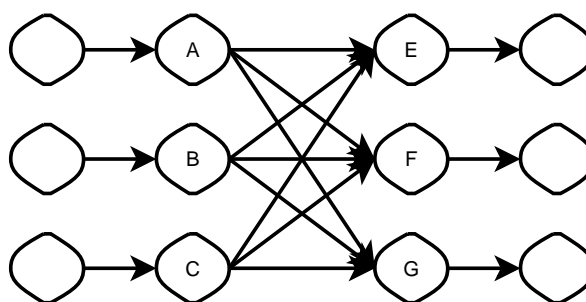


Abbildung 8: BGP AS Pfad

## 5.3 Verwandte AS

Es gibt Provider, die mehr als ein AS besitzen. Diese geben als AS Pfad häufig nur ein AS an, jedoch wird intern über Geräte geroutet, die Adressen haben, die unterschiedlichen AS zugeordnet sind.

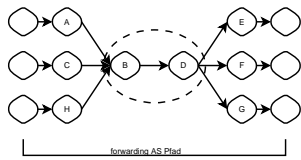
Wie man auch in Abbildung 9 erkennen kann, sind solche verwandten AS eine Ursache für zusätzliche AS Hops im forwarding Pfad. Da der forwarding Pfad auch über Geräte gehen kann, die wechselseitig einem der beiden AS zugeordnet sind, kann es dabei auch zu scheinbaren AS Schleifen kommen.

Es wird zur Identifizierung solcher verwandten AS ein ähnlicher Algorithmus angewandt, wie zur Erkennung von IXPs, nur wird diesmal ein Paar aus zwei AS betrachtet. Damit kann dann eine Liste von verwandten AS gepflegt werden, diese AS können dann bei der Betrachtung der Pfade als äquivalent angesehen werden.

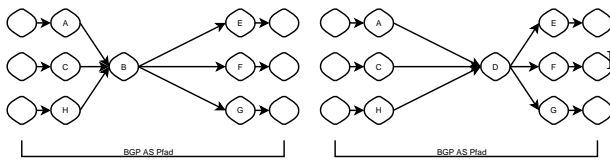
## 5.4 Adressen ohne BGP Eintrag

Solche Adressen werden für Router verwendet, da es nicht wichtig ist, dass der Router selbst erreicht werden kann, es reicht aus, wenn der Nachbarrouter weiß,

Abbildung 9: Verwandte AS



(a) forwarding Pfad



(b) BGP Pfad

(c) BGP Pfad

wie er ihn erreicht. Das kann z.B. mit einem internen Routing Protokoll sichergestellt werden. Trotzdem wird der Router natürlich in forwarding Pfaden auftreten. Wir nehmen das Beispiel aus Abbildung 10. AS C hat 2 Upstream Provider, A und B. Der Adressraum von C ist aus dem Adressraum von A alloziert. AS C benutzt nun einen Teil des Adressraums um seine Router zu adressieren, announced diesen Teil jedoch nicht. Auf Grund des kurzen (allgemeineren) Prefixes von A, das auch den Adressbereich von C umfasst, erscheinen diese Adressen nun AS A zugehörig. Folgende Fälle sind dadurch erklärbar:

**Extra Hop** Pfad 1 durchquert nur AS C, jedoch scheint AS A im forwarding Pfad aufzutauchen.

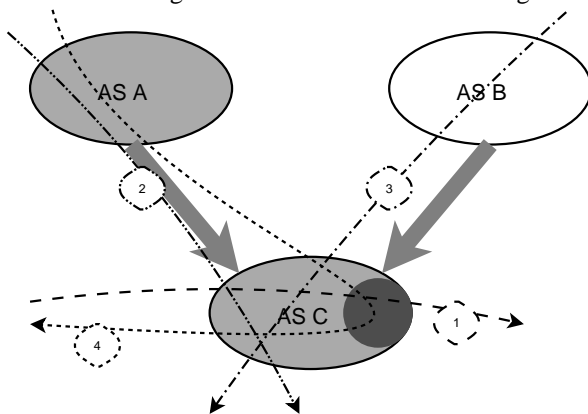
**Fehlender Hop** Pfad 2 durchquert AS C, jedoch scheint AS C im forwarding Pfad zu fehlen.

**Abweichender Hop** Pfad 3 läuft durch die AS B und C, jedoch scheint der forwarding Pfad durch B und A zu laufen.

**AS Schleife** Pfad 4 geht über A und C, der forwarding Pfad scheint aber die AS Route A C A zu haben

Wenn man sich auf die AS Schleifen konzentriert, dann stellt man fest, dass eine relativ kleine Zahl von fixen AS Paaren für die meisten solcher scheinbaren Schleifen verantwortlich ist. Man kann eine Liste solcher Paare pflegen, und immer, wenn ein Pfad durch AS C geht, dann werden zusätzliche Hops nach AS A trotzdem AS C zugeordnet.

Abbildung 10: Adressen ohne BGP Eintrag

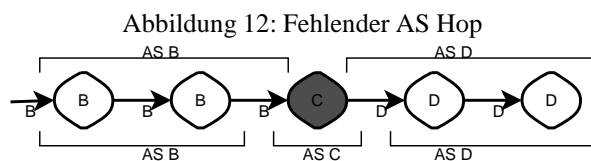
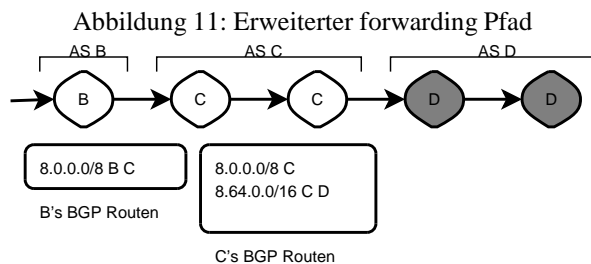


## 5.5 Ergebnis

Durch Anwendung dieser Methoden kann die Anzahl der nicht übereinstimmenden AS Pfade auf weit unter 5% gedrückt werden. Im folgenden werden einige der möglichen Ursachen erläutert.

## 6 Tatsächlich nicht übereinstimmende Pfade

In diesem Kapitel soll ein Versuch gemacht werden, zu erklären, warum BGP und forwarding Pfad manchmal tatsächlich nicht übereinstimmen.



## 6.1 Filter und Aggregate

An den Messpunkten steht immer nur eine vorgefilterte BGP Tabelle zur Verfügung. Dies kommt daher, dass AS unter Umständen einzelne Routen gefiltert haben, und insbesondere daher, dass Routen bereits aggregiert wurden, und daher die einzelnen AS nicht mehr unterscheidbar sind.

In Abbildung 11 kann man erkennen, wie Aggregation zu einem verlängerten forwarding Pfad führt. Der AS Pfad scheint bei AS C zu enden, obwohl der forwarding Pfad noch 2 Adressen, die AS D zugeordnet sind, liefert.

## 6.2 IP Adressen an AS Grenzen

Zwei Router stehen meist über eine Leitung mit einem lokalen Netzwerk in Verbindung. Daher haben beide Routerinterfaces Adressen aus dem selben Netz. Dieses ist typischerweise *einem* der beiden Provider zugeordnet, obwohl die Router *zwei* verschiedenen Providern gehören.

In Abbildung 12 kann man ein Beispiel sehen, bei dem AS C im BGP Pfad auftaucht, jedoch nicht im forwarding Pfad, da die beiden Link Netze jeweils den Providern B und D zugeordnet sind.

## 6.3 ICMP Quelladressen

Es ist nicht klar festgelegt, welches Interface, und daher auch welche Quelladresse, ein Router benutzen muss, um ein ICMP Paket zu versenden. Für diese Anwendung wäre es optimal, wenn alle Router das

Interface verwenden würden, auf dem sie das Testpaket empfangen haben. Meist spielt es zwar keine Rolle, da das ICMP Paket den selben Weg zurück nimmt, den auch das Testpaket genommen hat. Manchmal kann aber der Router abhängig von seiner Routingtabelle entscheiden das ICMP Paket in eine andere Richtung zu schicken. Dann verlässt das ICMP Paket den Router auf einem anderen Interface als dem, auf dem das Testpaket empfangen wurde. In diesem Fall macht es einen Unterschied, welche der beiden IP Adressen der Router als Quelladresse für das ICMP Paket wählt.

## 6.4 Routing Anomalien

Traceroute Pakete nehmen nicht unbedingt alle den gleichen Pfad. Traceroute geht aber genau davon aus. Wenn sich also das Routing während eines traceroute Laufs ändert, so sind die Ergebnisse eigentlich unbrauchbar und können zu verschiedensten Schlüssen führen. Es ist aber nicht nachprüfbar, ob das geschehen ist.

Außerdem können insbesondere Einflüsse von internen Routingprotokollen dazu führen, dass forwarding Pfad und BGP Pfad von einander abweichen, obwohl sowohl BGP als auch das interne Routingprotokoll stabil sind. Auch in diesem Fall sind die Ergebnisse nicht genau vorherzusagen.

## 7 Ergebnisse

In dieser Arbeit wurden verschiedene Methoden zur Pfadbestimmung im Internet vorgestellt und miteinander verglichen. Es konnte gezeigt werden, dass die üblichen Werkzeuge erhebliche Diskrepanzen in ihren Ergebnissen haben. Es wurden Methoden aufgezeigt, wie diese Diskrepanzen bis auf ein Minimum reduziert werden können, so dass eine zuverlässige Pfadbestimmung möglich wird. Abweichende oder unerwartete Ergebnisse können nun mit größerer Sicherheit als Indikator für eine tatsächliche Anomalie herangezogen werden.