Routing-Instabilitäten

Daniel Weber
(daniel.weber@in.tum.de)

Seminar „Internetrouting“,
Technische Universität München

WS 2004/2005 (Version vom 5. Februar 2005)

Zusammenfassung

Dieser Text betrachtet Verfahren zur Lokalisierung von Routing-Instabilitäten im Internet. Begonnen wird mit einem idealisierten Verfahren, das in der Praxis jedoch einigen Einschränkungen unterliegt. Deshalb wird eine Methode erarbeitet, die Informationen über verschiedene Präfixe, die von verschiedenen Beobachtungspunkten gesammelt wurden, kombiniert. Hierdurch sind die Ausgangspunkte von Instabilitäten relativ gut zu lokalisieren.

1 Einleitung

Die Nutzbarkeit des Internets hängt wesentlich von der Verfügbarkeit der Übertragungsnetze der verschiedenen Anbieter ab. Durch Instabilitäten im Routing, die z.B. aus Hardware- oder Softwarefehlern resultieren können, werden diese Übertragungswege in ihrer Leistungsfähigkeit beeinträchtigt oder gänzlich unbrauchbar gemacht. Zur Analyse solcher Störungen sowie zur Reduzierung ihrer Auswirkungen ist als erster Schritt eine möglichst exakte Bestimmung der auslösenden Stelle (des auslösenden autonomen Systems, AS) in der Netzwerktopologie notwendig.

Obwohl ein Ereignis – z.B. der Ausfall einer Leitung – dazu führt, dass die Information über die Wege zu den betroffenen Zielen aktualisiert werden, also diese Ziele über – wenn vorhanden – alternative Wege erreicht werden können, fehlt bei diesen Aktualisierungen die konkrete Information über die genaue, vom Ausfall betroffene Stelle. Diese Stelle muss mittels Korrelation der verschiedenen, durch den Ausfall ausgelösten, Veränderungen im Routing rekonstruiert bzw. eingegrenzt werden.

Hierzu wurden in den vergangenen Jahren verschiedenste Ansätze vorgestellt, die jedoch größtenteils an kleineren Schwächen leiden, die in bestimmten Konstellationen zu Tage treten. Am Beispiel zweier verwandter Verfahren soll hier der Weg von einem einfachen Ansatz zur Lokalisierung von Routing-Störungen zu einer praxistauglichen Bestimmung des Ausgangspunktes von BGP[1]-Störungen vorgestellt werden.

Zuerst wird ein idealisiertes Verfahren erläutert, das beim Vergleich von AS-Pfaden

[1] Border Gateway Protocol Version 4, siehe hierzu [Stew 99]

instabiler Präfixe[2] vor und nach einem Ereignis ansetzt. Diese AS-Pfade werden hierbei von sogenannten Beobachtungspunkten[3] ermittelt. Zusätzliche Informationen werden durch die Betrachtung mehrerer Präfixe auf einmal gewonnen. Das idealisierte Verfahren zeigt jedoch auch Einschränkungen bezüglich der Zuordenbarkeit von BGP-Updates, dem Auftreten gleichzeitiger Instabilitäten, der Dauer von Instabilitäten, der Erkennbarkeit von stabilen und besten AS-Pfaden und weiterem. Anhand dieser Einschränkungen und weiterer Nachteile werden Szenarien demonstriert, in denen das Verfahren keine korrekten Ergebnisse liefern kann.

Im nächsten Schritt wird ein angepasstes Verfahren beschrieben, dass geeignet ist, die Schwächen aufzuheben bzw. deren Auswirkung zu reduzieren. Das Verfahren befasst sich zuerst mit einer Methode zur Erkennung zusammengehörige BGP-Update-Bursts. Darauf basierend können dann die stabilen AS-Pfade vor und nach den untersuchten Ereignissen ermittelt werden. Im weiteren Verlauf wirde die Bildung der Kandidatenmengen aus den AS-Pfaden erläutert, wobei drei möglichen Methoden beschrieben wurden. Aus den Kandidatenmengen erfolgt dann die Ermittlung der endgültigen Kandidaten.

Abschließend wird die Umsetzung des angepassten Verfahrens in der Praxis geschildert. Dabei erfolgt zuerst die Erzeugung der Kandidatenmengen, die Zuordnung der auslösenden Ereignisse und schließlich die Anwendung einer Greedy-Heuristik um die am häufigsten betroffenen Kandidatentupel aus den Kandidatenmengen zu ermitteln. Dies wird zusätzlich mit Code-Beispielen veranschaulicht.

1.1 Routing im Internet

Die Grundlagen des Routings im Internet sollten bereits aus [Stew 99] oder anderen Quellen bekannt sein. Dennoch sollen diese im folgenden nochmals kurz in Erinnerung gerufen werden. Das Internet besteht aus einem Zusammenschluss vieler sogenannter autonomer Systeme (AS). Ein AS ist dabei ein Netz, das unter einer einheitlichen technischen und wirtschaftlichen Verwaltung steht und in der Regel das Netz eines Providers[4] darstellt.

Zwischen zwei verbundenen autonomen Systemen treten dabei vor allem zwei Beziehungen auf:

Peerings:
Hierbei tauschen die beteiligten autonomen Systeme den Traffic für das jeweils andere AS sowie dessen Kunden direkt und kostenneutral aus.

Upstreams:
Dabei kauft ein AS von einem anderen AS, dem Upstream, Konnektivität ein um den „Rest“ des Internets darüber zu erreichen.

Die Informationen über die erreichbaren Präfixe werden zwischen den beteiligten autonomen Systemen mittels BGP ausgetauscht. Diese Informationen werden von AS zu AS propagiert und ermöglichen den Anbietern im Idealfall den Aufbau einer vollständigen

[2]Ein Präfix ist ein Teil der Routing-Information, die notwendig ist, um einen bestimmten IP-Adress-Bereich zu erreichen. Alle IP-Adressen in diesem Bereich haben das gleiche Präfix. Die Angabe „131.159/16“ bezeichnet das Präfix der Fakultät Mathematik Informatik der Technischen Universität München.

[3]Beobachtungspunkte werden in einigen ASen aufgestellt und nehmen dort passiv an der BGP-Kommunikation teil. D.h. sie protokollieren alle BGP-Updates aus ihrer jeweiligen Perspektive für eine spätere Auswertung mit.

[4]Große internationale Provider unterteilen ihre Netze manchmal auch in mehrere ASe für unterschiedliche Regionen oder Länder.

Tabelle an Präfixen, dem Weg zu diesen Präfixen und möglicherweise weiteren Attributen.

Hinweise auf Instabilitäten erhält man durch Veränderungen, die sich im laufenden Betrieb in diesen Tabellen ergeben. Dabei können diese Instabilitäten entweder zwischen zwei Anbietern — auf einer AS-AS-Kante — oder innerhalb eines AS auftreten.

2 Lokalisierung von Instabilitäten im Idealfall

In diesem Abschnitt geht es um das Idealbild einer Methode zur Lokalisierung von Routing-Instabilitäten. Jede dieser Instabilitäten führt zu einer Reihe von BGP-Updates, die von den Routern direkt neben der Instabilität durch das Internet propagiert werden. An den Beobachtungspunkten werden diese BGP-Updates für die spätere Auswertung und damit Lokalisierung der Instabilität aufgezeichnet. Die Instabilitäten selbst können hierbei entweder auf der Verbindung zwischen zweier ASe oder innerhalb eines AS stattfinden. Im ersten Fall befindet sich die Störung auf einer AS-AS-Kante, im zweiten Fall befindet sie sich in einem AS.

Die Grundidee der Lokalisierung liegt darin, dass eine Instabilität, die zu einer Änderung des besten AS-Pfades für ein Präfix durch BGP-Updates führt, aus einem Ereignis entsteht, das entweder auf dem alten oder auf dem neuen AS-Pfad stattgefunden hat. Es ist anzunehmen, dass das Ereignis auf dem besseren AS-Pfad[5] zwischen Präfix und Beobachtungspunkt stattgefunden hat, denn in diesem Fall kann das Ereignis eine Änderung vom alten auf den neuen besten AS-Pfad verursachen. Je nachdem, ob der AS-Pfad vor oder nach dem Ereignis besser war, ist auch erkennbar, ob es sich bei dem Ereignis um z.B. den Anfang oder das Ende einer Störung handelt. Nur durch den Anfang einer Störung wird auf einen schlechteren AS-Pfad gewechselt. Steht nach dem Ende der Störung der alte, bessere AS-Pfad wieder zur Verfügung, wird wieder auf ihn zurückgeschwenkt. Da nun von außerhalb möglicherweise nicht erkennbar ist, welcher der beiden AS-Pfade der bessere[6] ist, ist es zweckmäßig, die Vereinigungsmenge beider AS-Pfade als Kandidatenmenge für die Störungsursache anzusehen.

Durch eine Ausweitung der Perspektive auf mehrere Präfixe lässt sich die mögliche Quelle der Instabilität noch weiter eingrenzen. Betrachtet man eine einzige Instabilität innerhalb einer bestimmten Zeitspanne, so ist klar, dass alle AS-Pfad-Änderungen für verschiedenste Präfixe ihre Ursache in dieser Instabilität haben müssen. Jede dieser AS-Pfad-Änderungen ist nun an verschiedenen Beobachtungspunkten wahrnehmbar; daraus folgt wiederum, dass die mögliche Quelle der Instabilität innerhalb der Schnittmenge der verschiedenen ermittelten AS-Pfade der jeweiligen Präfixe und Beobachtungspunkte liegen muss. Durch das Fehlen von AS-Pfad-Änderungen für andere Präfixe und/oder andere Beobachtungspunkte erhält man eine zusätzliche Information: Über den beobachteten Zeitraum muss der AS-Pfad zwischen einem Präfix und dem Beobachtungspunkt stabil gewesen sein. Die Kanten dieser stabilen AS-Pfade können damit als mögliche Quelle der Instabilität ausgeschlossen werden.

Dieses idealisierte Verfahren unterliegt jedoch einigen einschränkenden Anforderungen:

- Allen BGP-Updates lässt muss die auslösende Instabilität zugeordnet werden können.

[5] Der bessere AS-Pfad ist in erster Linie der kürzere AS-Pfad. Es gibt zwar weitere Attribute, die ebenfalls entscheidungsrelevant sein können (siehe [Stew 99]), diese sind jedoch nicht verlässlich über AS-Grenzen hinweg bekannt.

[6] Stewart führt in [Stew 99] bis zu sechs Schritte zur Auswahl auf; hier steht einzig die AS-Pfad-Länge als Kriterium zur Verfügung.

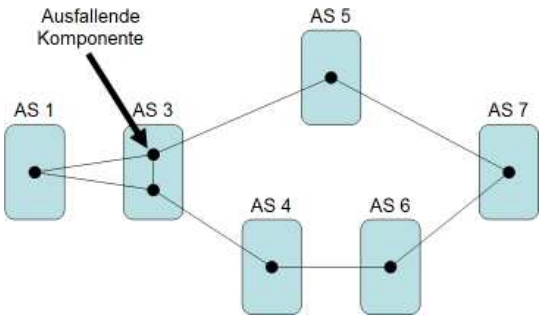

Abbildung 1: gemeinsame Abschnitte

- Jedes Präfix ist jederzeit von höchstens einer Instabilität betroffen.
- Die Zeitspanne für den BGP-Konvergenzprozess [LAWV 01] ist endlich und festgelegt.
- Es ist klar erkennbar, welche AS-Pfade stabil sind.
- Es ist ebenfalls klar erkennbar, welcher AS-Pfad von zweien der bessere ist.
- Es treten keine induzierten Instabilitäten [FM 04] auf.

2.1 Aber Vorsicht...

...die Annahmen, die dieser Idealfall vorraussetzt, sind im realen Betrieb beim BGP-Routing nicht verlässlich gegeben. In den folgenden Beispielen wird an Einzelfällen gezeigt, warum diese Annahmen für die Lokalisierung von Instabilitäten zweckmäßig sind (detailierter wird darauf in [TR 04] eingegangen). Dennoch ist es möglich, das Verfahren hierauf anzupassen. Dies wird im im nächsten Abschnitt erläutert.

2.1.1 Risiko: Ausschluss von Störungsquellen

Sieht man von einem Beobachtungspunkt aus für ein bestimmtes Präfix zuerst einen AS-Pfad 1-3-5-7 und nach einem Ereignis den AS-Pfad 1-3-4-6-7, wie in Abbildung 1, so könnte man daraus schließen, dass die gemeinsamen Abschnitte des alten und neuen AS-Pfades, also 1-3 und 7, nicht als Ursache der Störung in Frage kommen [CGH 03].

Wie man jedoch im Beispiel erkennen kann, wurde die Instabilität durch den Ausfall des markierten Routers innerhalb des AS 3 ausgelöst, also in gerade einem der ASe, die aufgrund der gemeinsamen AS-Pfad-Abschnitte zwischen altem und neuem AS-Pfad als Störungsursache ausgeschlossen werden sollten.

Im realen Betrieb sind die Szenarien aber noch weitaus komplexer. Dort spielen möglicherweise IGP[7], MED[8], Communities und Local-Pref-Attribute mit hinein und lassen Ereignisse innerhalb eines AS sehr weit propagieren.

Es ist ist auch ein ähnliches Szenario denkbar, bei dem die Verbindung zwischen zwei ASen eine Instabilität erfährt, diese Instabilität aber nur begrenzt propagiert wird, siehe [FM 04]. Dadurch kann es von zwei unterschiedlichen Beobachtungspunkten für ein-

[7]Internal Gateway Protocol, das jeweils innerhalb eines AS verwendete Routing-Protokoll, z.B. OSPF oder ISIS.

[8]Multi-Exit-Discriminator, eine Möglichkeit zur Bevorzugung bestimmter Verbindungen, z.B. falls zwei ASe an mehreren Orten miteinander verbunden sind. Siehe auch [Stew 99].

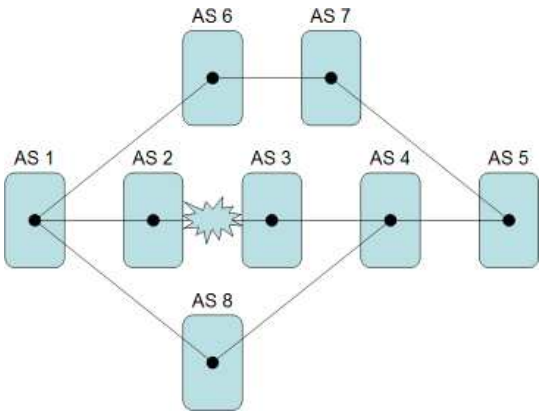

Abbildung 2: induzierte Störungen

und dasselbe Präfix einmal eine Änderung des AS-Pfades geben und einmal nicht, obwohl in beiden Fällen die AS-Pfade vor der Instabilität einen teilweise identischen Weg hatten. Die Tatsache, dass der eine Beobachtungspunkt für das Präfix keine Änderung des AS-Pfades „gesehen“ hat, könnte man nun nutzen, um die ASe auf diesem AS-Pfad als Störungsquelle auszuschließen. Dabei würde in diesem Fall allerdings auch die eigentliche Quelle der Instabilität fälschlicherweise ausgeschlossen werden.

2.1.2 Risiko: Induzierte Updates

Betrachtet man Abbildung 2, so wird ebenfalls klar, dass die Ursache für eine Instabilität doch nicht zwingenderweise auf dem alten oder dem neuen AS-Pfad liegen muss.

Ein Beobachtungspunkt, der am AS 5 angeschlossen ist, sieht als AS-Pfad zu einem im AS 1 liegenden Präfix den kürzesten AS-Pfad 1-6-7-5. Im AS 4 wurde eine Bevorzugung des AS-Pfades 1-2-3-4 statt des kürzeren AS-Pfades 1-8-4 konfiguriert, d.h. der AS-Pfad 1-2-3-4 wird auch an die Nachbarn, also auch an AS 5, propagiert. Jedoch bevorzugt AS 5 weiterhin die kürzere Route. Fällt nun z.B. die Verbindung zwischen AS 2 und 3 aus, so ist der im AS 4 verwendete AS-Pfad nicht mehr verfügbar, d.h. AS 4 wechselt auf den bisher unterdrückten AS-Pfad 1-8-4. Dadurch ändert sich auch das Announcement[9] an AS 5, wodurch dort nun möglicherweise der bevorzugte AS-Pfad auf 1-8-4-5 wechselt.

Wie man erkennt, ist der tatsächlich gestörte Abschnitt weder im alten AS-Pfad 1-6-7-5 noch im neuen AS-Pfad 1-8-4-5 enthalten. Die Ursache der Instabilität nur auf dem alten oder neuen AS-Pfad zu suchen kann also auch zu falschen Ergebnissen führen.

3 Das angepasste Verfahren

Vor der Anwendung des eigentlichen Verfahrens sind noch einige Vorbereitungen notwendig, um kleinere Problemstellungen aus dem Verfahren herauszunehmen. Zum einen geht es dabei um ein geeignetes Verfahren zur Erkennung zusammenhängender BGP-

[9]Ein Announcement ist ein BGP-Update an einen benachbarten BGP-Router, um diesem mitzuteilen, dass ein bestimmtes Präfix über den Sender des Announcements erreichbar ist oder diesem zumindest der nächste Hop [Stew 99] bekannt ist.

Update-Bursts[10], zum anderen um eine Methode zum ersten Ausschluss mancher ASe oder AS-AS-Verbindungen aus der Menge der Kandidaten.

3.1 Bestimmung der AS-Pfade vor und nach Ereignissen

Bisher wurde davon ausgegangen, dass die stabilen AS-Pfade vor und nach einem auslösenden Ereignis bereits vorliegen. Folglich könnten diese sofort für die Lokalisierung des Ereignisses verwendet werden. Da die Ermittlung dieser stabilen AS-Pfade jedoch nicht derart trivial ist, muss deren Bestimmung vorher noch etwas zusätzliche Aufmerksamkeit gewidmet werden.

In der Praxis hat man den AS-Pfad vor einem Ereignis vorliegen, muss jedoch die BGP-Konvergenzzeit abwarten, um den stabilen AS-Pfad nach einem Ereignis zu kennen. Ursache dafür ist, dass auf ein Ereignis nicht nur ein einzelnes BGP-Update erfolgt, sondern über eine längere Zeitspanne Update-Bursts eingehen können. Erst wenn die Route auf einen stabilen AS-Pfad konvergiert ist, enden die Update-Bursts.

Um den stabilen AS-Pfad nach dem Ereignis ermitteln zu können, bedarf es einer möglichst akkuraten Erkennung der zum Ereignis gehörenden BGP-Update-Bursts. Hierzu stehen verschiedene Timeouts[11] zur Verfügung:

Relativer Timeout:
Als Timeout wird die Zeitspanne zwischen dem Anfang des ersten BGP-Update-Bursts und dessen Ende verwendet. Nun wird die Ruhephase gemessen, die vor dem Eingehen der folgende BGP-Update-Bursts herrscht. Ist diese Ruhephase kürzer als die Timeout-Zeitspanne, dann wird der nachfolgende BGP-Update-Burst dem Ereignis zugeordnet. Ist die Ruhepause länger als die Timeout-Zeitspanne, wird von einem neuen Ereignis ausgegangen. Der Nachteil dieses Timeouts ist, dass verschiedene tatsächliche Ereignisse über eine lange Zeitspanne fälschlicherweise dem gleichen Ereignis zugeordnet werden können.

Statischer Timeout:
Hier wird eine feste Timeout-Zeitspanne verwendet, wodurch man jedoch Empfindlichkeit einbüßt. Wie in [FM 04] erwähnt, kann ein für ein Präfix geeigneter Timeout bei einem anderen Präfix gänzlich ungeeignet sein; so können z.B. zusammengehörige BGP-Update-Bursts unterschiedlichen Ereignissen zugeordnet werden.

Adaptiver Timeout:
Beim adaptiven Timeout wird die Erkennung zusammengehöriger BGP-Update-Bursts in zwei Abschnitte unterteilt. Zuerst wird über eine Zeitspanne t ein relativer Timeout von $t/2$ verwendet. Ist die Zeitspanne t abgelaufen und wurde noch kein stabiler AS-Pfad erreicht, greift ein relativer Timeout von 0. D.h. dass das Ende von aktuell noch laufenden BGP-Update-Bursts abgewartet und dann sofort unterbrochen wird, ohne noch eine Zeitspanne abzuwarten.

[10] Durch ein Ereignis enstehen in der Regel mehrere BGP-Updates, zum einen direkt von der betroffenen Stelle ausgehend, zum anderen von den Empfängern dieser BGP-Updates u.s.w. induzierte, weitere Mitteilungen. Diese Folge von zusammengehörigen BGP-Updates wird BGP-Update-Burst genannt.

[11] Timeout bezeichnet allgemein eine Zeitspanne der Inaktivität nach deren Ablauf vom endgültige Ende einer vorherigen Aktivität ausgegangen wird. In diesem speziellen Fall wird nach dem Ablauf des Timeouts vom endgültigen Ende der vorherigen BGP-Update-Bursts ausgegangen.

Aufgrund der Nachteile des relativen und statischen Timeouts wird die Methode mit adaptivem Timeout verwendet. Zusammen mit dem Ausschluss *flappender*[12] Updates werden die Ereignisse in einem zeitlich akzeptablen Rahmen gehalten und auch parallele BGP-Update-Bursts korrekt zugeordnet.

3.2 Bildung der Kandidatenmengen bei Ereignissen

Anhand der AS-Pfade vor und nach einem Ereignis, die mit dem bereits in Abschnitt 3.1 genannten adaptiven Timeout aus den BGP-Update-Bursts ermittelt wurden, sollen die wahrscheinlichsten ereignisauslösenden Kandidaten eingeschränkt werden.

Eine erste denkbare Methode, die in [FM 04] und [CGH 04] erwähnt wird, ist der Ausschluß gemeinsamer Abschnitte im AS-Pfad (*shared path segments*). Betrachtet man z.B. den alten AS-Pfad 1-3-5-7 und nach dem Ereignis den neuen AS-Pfad 1-3-4-6-7 so ist die Versuchung groß, AS 5 oder die Verbindungen 3-5 oder 5-7 als Ereignisauslöser zu vermuten, da die restlichen ASe in beiden AS-Pfaden enthalten sind.

Der Ausschluß von gemeinsamen Abschnitten zu Anfang oder Ende der AS-Pfade ist jedoch nicht zuverlässig anwendbar. So können in obigem Beispiel Ereignisse im internen Routing (*IGP*) von AS 3 oder 7 zur Veränderung des AS-Pfades führen. Somit ist diese Methode nicht für die Lokalisierung des Ereignisses geeignet.

Als weitere Methode kommt in [FM 04] der Ausschluß des schlechteren AS-Pfades zur Anwendung. Dies basiert darauf, dass davon auszugehen ist, dass das auslösende Ereignis generell auf dem besseren AS-Pfad stattgefunden hat. Im Falle einer Störung findet nur dann der Wechsel zum schlechteren Pfad statt, wenn der bessere AS-Pfad aufgrund der Störung nicht mehr zur Verfügung steht. Wird hingegen zu einem besseren AS-Pfad gewechselt, dann handelt es sich i.d.R. um das Ende einer Störung auf diesem besseren AS-Pfad. Folglich muss das jeweils betrachtete Ereignis dort stattgefunden haben.

Mittels der AS-Pfade vor und nach einem Ereignis, die mit dem bereits in Abschnitt 3.1 genannten adaptiven Timeout aus den Update-Bursts ermittelt werden, können nun die für das Ereignis verantwortlichen ASe oder AS-zu-AS-Verbindungen gebildet werden. Diese werden im folgenden als Kandidatentupel[13] bezeichnet.

3.3 Ermittlung der endgültigen Kandidaten

Da von einem Ereignis, z.B. dem Ausfall einer Leitung, i.d.R. nicht nur ein, sondern mehrere Präfixe betroffen sind, können zusätzliche Informationen abgeleitet werden. Haben zwei von einem Ereignis betroffene Präfixe eine nicht-leere Schnittmenge aus deren Kandidatenmengen, lässt sich daraus ableiten, dass die Ursache des Ereignisses in dieser Schnittmenge zu vermuten ist.

[12] Bei „flappenden Routen“ handelt es sich um Präfixe, die immer wieder veröffentlicht und zurückgezogen werden. Um die Stabilität des Internet-Routings zu erhöhen haben einige Router-Hersteller Dämpfungs-Algorithmen gegen zu häufige Flaps entwickelt.

[13] Kandidatentupel bestehen aus Paaren von AS-Nummern (n, m), wobei ein solches Tupel mit $n \neq m$ für ein mögliches Ereignis entlang der Verbindung zwischen AS n und AS m steht. Ein Tupel mit $n = m$ steht hingegen für ein mögliches Ereignis innerhalb des AS n.

4 Das eigentliche Verfahren

Nachdem die Vorarbeit geleistet ist, lassen sich jetzt die einzelnen Bestandteile zum eigentlichen Verfahren zusammenfügen. Folgende Schritte werden hier der Reihe nach ausgeführt:

- Erzeugen der Kandidatenmengen
- Ermitteln der zugeordneten Ereignisse
- Greedy-Heuristik zur Ermittlung der am häufigsten betroffenen Kandidatentupel

4.1 Erzeugen der Kandidatenmengen

Im ersten Schritt sammeln wir für jedes Präfix aus der Präfixmenge und für jeden Beobachtungspunkt die jeweiligen BGP-Updates abzüglich der BGP-Updates, bei denen es sich lediglich um Route-Flaps gehandelt hat. Diese BGP-Updates werden zu BGP-Update-Bursts gruppiert und mittels des oben beschriebenen adaptiven Timeouts wird aus ihnen der stabile AS-Pfad vor und nach dem auslösenden Ereignis erzeugt. Aus den beiden ermittelten AS-Pfaden erfolgt die Auswahl des besten AS-Pfades, der nun zur Bildung der Menge der Kandidatentupel genutzt wird.

```
for each prefix p
{
  for each observationpoint o
  {
    U := updates_p( o ) − flaps_p( o )
    B := updateburst( U, timeout )

    for each burst b in B
    {
      r_o := as_path( old_stable_route( b ) )
      r_n := as_path( new_stable_route( b ) )
      r_b := best_as_path( r_o, r_n )
      candidate_set c_ob := candidates( r_b )
    }
  }
}
```

4.2 Ermitteln der Ereignisse

Anhand der bei den Beobachtungspunkten eingegangenen BGP-Update-Bursts werden nun alle für die Beobachtung relevanten Ereignisse ermittelt. Jedem Ereignis werden die dadurch ausgelösten BGP-Update-Bursts zugeordnet.

```
for each timeunit t
{
  for each prefix p
  {
    E_p := E_p ∪ new_event( t )
  }
}

for each event e in E_p
{
  event_burst_set_e := associate_event_bursts( B, e )
```

}

4.3 Vereinigung der Kandidatenmengen

Für jedes der erkannten Ereignisse, jedes beteiligte Präfix und alle Beobachtungspunkte werden anhand der zugehörigen BGP-Update-Bursts die oben bereits erzeugten Kandidatenmengen je Beobachtungspunkt vereinigt. Die leeren Kandidatensets werden für jeden Beobachtungspunkt gesondert betrachtet. Dabei treten keine Änderungen im AS-Pfad auf, somit hat für die jeweiligen Präfixe kein Ereignis auf diesem Weg stattgefunden, was im folgenden genutzt wird:

Aus dem stabilen AS-Pfad der Präfixe, für die keine Änderungen stattgefunden haben, werden dennoch Kandidatenmengen gebildet. Da jedoch bekannt ist, dass diese keine Instabilitäten aufweisen, werden diese Kandidatenmengen von der obigen Vereinigung aller Kandidatenmengen abgezogen. Es werden also stabile Kandidatentupel ausgeschlossen.

```
for each event e
{
  for each prefix p
  {
    for each observationpoint o
    {
      for each burst (b, o) in event_burst_set_e
      {
        candidate_set c_o := ⋃ c_ob
      }
    }
    for each observationpoint o
    {
      if candidate_set c_o == ∅
      {
        candidate_set s_o := stable_route( o, e )
      }
    }
    instability_candidates := ⋂ c_o − ⋃ s_o
  }
}
```

4.4 Greedy-Heuristik

Nun werden die über die beobachtete Zeitspanne auf den einzelnen Beobachtungspunkten bemerkten Ereignisse einander zugeordnet und daraus die Menge der korrelierten Ereignisse gebildet. Für jedes der korrelierten Ereignisse werden wiederum die einzelnen beobachteten Ereignisse zusammengestellt und anhand dieser für das jeweilige korrelierte Ereignis alle davon betroffenen Präfixe ermittelt.

Solange die Menge der betroffenen Präfixe nicht leer ist, wird daraus ein Präfix ausgewählt. Anhand dieses Präfix wird das auslösende Ereignis genutzt, um die Kandidatenmengen aller betroffenen Präfixe zu ermitteln. In diesen Kandidatenmengen werden nun identische Tupel[14] gezählt. Das am hierbei am häufigsten vorkommende Tupel wird

[14]Mit identischen Tupeln sind Tupel gemeint, die die gleichen AS-Nummern enthalten, unabhängig von der Reihenfolge. (n, m) wird also als identisch zu (m, n) angesehen.

als mögliche Instabilität vermerkt. Alle von dieser Instabilität betroffenen Präfixe werden dann aus der Präfixmenge entfernt.

Nach Abschluß hat man die Menge aller möglichen Instabilitäten aus den Kandidatenmengen extrahiert.

5 Bewertung des Verfahrens

Zur Beurteilung der Praxistauglichkeit des Verfahrens wurden BGP-Protokolle und BGP-Tabellen aus dem Zeitraum vom 4. bis zum 16. Dezember 2003[15] gesammelt. Aus diesem Material wurde die AS-Topologie als Graph mit 16.757 Knoten und 45.376 Kanten abgebildet. Dabei enthalten waren 30.653 Kunden-Anbieter-Beziehungen und 1.532 Peering-Beziehungen. Auf diesen Daten basierend wurde nun durch Simulationen die Anwendbarkeit geprüft.

Bei der Lokalisierung der Quellen der simulierten Instabilitäten traten folgende interessante Aspekte zu Tage:

- Die Stelle einer Instabilität hat Einfluß auf die Lokalisierbarkeit der Instabilität. Instabilitäten auf sogenannten „top tier"-Kanten[16] sind bei 69% der Beobachtungspunkte sichtbar, bei „middle tier"-Kanten beläuft sich dies auf fast 40% und bei „bottom tier"-Kanten etwa 15%.
- Mit der Standard-Heuristik und nur zwei Beobachtungspunkten lassen sich bereits mehr als 68% aller Instabilitäten auf 5 bis 7 ursächliche Kanten eingrenzen. Mit 10 Beobachtungspunkten steigt dies auf nahezu 88% an.
- Kombiniert man alle Heuristiken – wie im Verfahren geschehen – kann für mehr als 88% aller Instabilitäten die Ursache auf weniger als 5 Kanten eingegrenzt werden.

6 Zusammenfassung und Ausblick

Wie man auf den vergangenen Seiten gesehen hat, stellt sich die Lokalisierung von Routing-Instabilitäten als nicht-triviales Problem dar. Dies lässt sich dadurch erklären, dass die Publizierung der Störungsstelle im Border Gateway Protocol nicht vorgesehen ist. Die fehlende Information muss also durch die Korrelation anderer, im BGP noch vorhandener, Informationen wiederhergestellt werden. Eine weitere Einschränkung stellt die begrenzte Sicht auf das Internet dar. Sämtliche zur Verfügung stehenden Informationen müssen an dedizierten Beobachtungspunkten, welche passiv am BGP teilnehmen, gewonnen werden. Dennoch zeigt das Schrittweise aufgebaute Verfahren mit der Möglichkeit, 88% aller Instabilitäten auf vier oder weniger AS-Kanten einzugrenzen, seine Leistungsfähigkeit bei der Suche der Störungsstelle.

Das Potential für weitere Arbeiten in diesem Bereich ist dennoch nicht erschöpft. Denkbar ist eine detailiertere Untersuchung der Ausbreitung von Routing-Instabilitäten durch BGP. Ebenfalls von Interesse wäre ein möglicher Zusammenhang zwischen der Ausbreitung einer Routing-Instabilität und der Entfernung von deren Quelle zum „Zentrum" des Internets. Analog bietet könnte analysiert werden, welchen Einfluss die Standorte der

[15] Die Details zur Zusammensetzung der gesammelten Daten können [FM 04] entnommen werden.

[16] Netzbetreiber werden üblicherweise in eine Tier-Hierarchie eingeordnet. Zum Beispiel kann ein Tier-1-Provider seinen sämtlichen Datenverkehr über Peerings und Kunden abwickeln, er benötigt keinerlei Upstreams für den Datenverkehr. Dabei sind „top tier"-Kanten all jene, an denen ein Tier-1-Provider beteiligt ist, „middle tier"-Kanten sind die restlichen, an denen ein Tier-2-Provider beteiligt ist und „bottom tier"-Kanten sind alle verbleibenden.

Beobachtungspunkte – wieder relativ zum „Zentrum“ des Internets – auf die Ergebnisse des Verfahrens haben.

Literatur

[FM 04] Anja Feldmann, Olaf Maennel, Z. Morley Mao, Arthur Berger, Bruce Maggs: *Locating Internet Routing Instabilities*; SIGCOMM 2004

[Stew 99] John W. Stewart III: *BGP4: Inter-Domain Routing in the Internet*; Addison Wesley Longman, Inc., 1999.

[CGH 04] Di-Fa Chang, Ramesh Govindan, John Heidemann: *Locating BGP Missing Routes Using Multiple Perspectives*; SIGCOMM 2004

[TR 04] R. Teixeira und J. Rexford: *A measurement framework for pin-pointing routing changes*; SIGCOMM 2004

[CGH 03] Di-Fa Chang, Ramesh Govindan, John Heidemann: *The temporal and topological characteristics of BGP path changes*; Proc. ICNP 2003

[LAWV 01] Craig Labovitz, Ahba Ahuja, Roger Wattenhofer, Srinivasan Venkatachatry: *The Impact of Internet Policy and Topology on Delayed Routing Convergence*; INFOCOM 2001