

Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen

Sebastian Feld – Tim Perrei – Norbert Pohlmann – Matthias Schupp

Institut für Internet-Sicherheit – if(is)
Fachhochschule Gelsenkirchen
{feld | perrei | pohlmann | schupp}@internet-sicherheit.de

Zusammenfassung

Das Internet ist eine kritische Infrastruktur, deren Verfügbarkeit für die Bürger, Unternehmen und Regierungen der Industrieländer von höchster Bedeutung ist. Um die Komplexität einer solchen Infrastruktur analysieren, den aktuellen Zustand beobachten sowie die zukünftige Entwicklung abschätzen zu können ist der Einsatz eines Kennzahlensystems, wie es bereits in der Betriebswirtschaft breite Anwendung findet, sinnvoll. Für ein solches Internet-Kennzahlensystem stellen sich drei Anforderungen: Es muss die Daten, welche die Kennzahlen bilden werden, erfassen, verarbeiten und visualisieren können. Bei der Realisierung eines solchen Internet-Kennzahlensystems ist insbesondere ein Augenmerk auf die Erhebung der Kennzahlen sinnvoll. Dabei kann zwischen Datenquellen, die ihre Daten selbstständig einliefern, und jenen, deren Daten abgerufen werden müssen, unterschieden werden. Auch wenn das Internet eine globale Infrastruktur ist, so ist ein Kennzahlensystem für ein „Internet Deutschland“ notwendig und sinnvoll. Dabei ist der Geltungsbereich insbesondere durch die Infrastruktur und Dienste gegeben, die direkt für die deutschen Bürger und Unternehmen relevant sind.

1 Die kritische Infrastruktur Internet

Das heutige Internet besteht aus einer Vielzahl von Netzwerken, sogenannte Autonome Systeme (AS), die von Internet Service Providern (ISP), großen Unternehmen oder Hochschulen betrieben werden. AS können entweder direkt miteinander (Private Peering), oder an öffentlichen Stellen (Public Peering) verbunden werden. Die Stellen für Public Peering werden auch als Internetknoten oder Internet Exchange Point (IXP) bezeichnet. Zurzeit gibt es mehr als 38.000 AS¹, die mit mehr als 70.000 Verbindungen das Internet bilden. Statistisch gesehen bleibt ca. 30% der Kommunikation im eigenen AS, 11% wird über Public- und 26% über Private-Peering abgewickelt. 33% der Kommunikation stellt sogenannten Transit (auch Upstream genannt) dar, für den bezahlt werden muss [Wott10].

Für eine genauere Betrachtung und eine richtige Einschätzung der Bedeutung von einzelnen AS ist es wichtig zu sehen, welche Rolle das jeweilige AS im Zusammenspiel des Internet-Verbundes einnimmt. Ein AS ist ein Netz aus Routern und Teilnetzen, und untersteht einer einzigen administrativen Instanz. Diese Netze, die sich in Größe und räumlicher Ausdehnung immens unterscheiden können, handeln absolut autonom. Das heißt, sie werden unterschiedlich und vollkommen unabhängig voneinander verwaltet. Das wiederum bedeutet, dass die

¹Vgl. unter anderem die Routing-Tabelle des Projekts „Route Views“ unter routeviews.org.

2 Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen
Betreiber beispielsweise unterschiedliche Strategien haben, wie sie mit Hilfe von Routing-Protokollen die Kommunikation der IP-Pakete in ihren Netzen organisieren.

Nicht nur innerhalb Deutschlands zählt das Internet als Teil der Informations- und Telekommunikationstechnologie zu den kritischen Infrastrukturen. Weitere technische Basisinfrastrukturen sind etwa die Energieversorgung, Transport und Verkehr sowie die Wasser- und Abwasserversorgung [BMI09]. Die Beeinträchtigung oder der Ausfall von Teilen des Internets kann enorme Auswirkungen haben. Beispielsweise kann die Beeinträchtigung der IP-Telefonie, wie beim Skype-Ausfall im Jahre 2007², zu nachhaltigen wirtschaftlichen Schäden führen, wenn ein Unternehmen nicht mehr in der Lage ist, telefonische Geschäfte abzuwickeln.

Gegenwärtig ist ein fortwährender Betrieb des Internets unabdingbar. Um einen möglichst störfreien Betrieb gewährleisten zu können ist es notwendig, den aktuellen Zustand des Internets einsehen, und außerdem die zukünftige Entwicklung desselben einschätzen zu können. Nur so kann auf neuartige Begebenheiten (sowohl positiver als auch negativer Natur) optimal eingegangen werden.

2 Motivation für ein Internet-Kennzahlensystem

Die Beschreibung der Motivation eines Internet-Kennzahlensystems setzt die Abgrenzung einiger Begrifflichkeiten voraus. Im Folgenden ist mit einer **Kennzahl** das Abbild eines Teils der Realität gemeint. Eine Kennzahl ist somit eine Zahl, die einen quantitativ erfassbaren Sachverhalt in konzentrierter Form wiedergibt [Kuet03]. Darauf aufbauend ist mit einem **Kennzahlensystem** die Kombination mehrerer Kennzahlen gemeint. Während eine Kennzahl zwar verschiedentlich interpretiert werden kann, sich jedoch nur auf ein bestimmtes Merkmal bezieht, kann ein Kennzahlensystem einen komplexeren Sachverhalt mit mehreren Merkmalen darstellen. Ein **Internet-Kennzahlensystem** ist schließlich ein Kennzahlensystem, bei dem der zu untersuchende Sachverhalt das Internet als solches ist. Es werden Kennzahlen gesammelt, die sich auf das Internet beziehen oder die durch das Internet generiert werden.

Grundsätzlich lässt sich die Motivation für ein Internet-Kennzahlensystem in zwei Aspekte einteilen. Zum einen geht es um die vielen unterschiedlichen Abhängigkeiten im Kontext des Internets, zum anderen um kaum vorhandene bereits bestehende Kennzahlen.

2.1 Abhängigkeiten im Kontext des Internets

Es gibt viele unterschiedliche Abhängigkeiten im Kontext des Internets, wobei in einigen Fällen eine Abhängigkeit von einem Staat (oftmals die Vereinigten Staaten von Amerika) festzustellen ist. Die Abhängigkeiten können auf der Ebene der Technik (zum Beispiel die Verbindung von Teilen des Internets mittels interkontinental verlegter Seekabel), auf der Ebene der Dienste (beispielsweise ist das „Websurfen“ ohne die transparente Verwendung des Domain Name System (DNS) kaum praktikabel) oder auf der Ebene der Verwaltung (zum Beispiel koordiniert die Internet Corporation for Assigned Names and Numbers (ICANN) unter anderem die Verwaltung der Top-Level-Domains) existieren. Im Folgenden seien zwei Beispiele für Abhängigkeiten im Kontext des Internets hervorgehoben.

Es existieren wenige sehr große AS (sogenannte Tier-1-Provider), die weite Teile des Internets miteinander verbinden und damit eine Konnektierung aller Endsysteme im ganzen Internet erreichen. Diese sind für die Stabilität des Internets enorm wichtig. Die größten und

²Vgl. <http://heartbeat.skype.com/2007/08/>

Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen 3
wichtigsten AS sind derzeit US-amerikanisch. Das größte deutsche AS, das der Deutschen Telekom, ist in den unteren Plätzen der TOP25 nach Anzahl der Verbindungen zu finden³.

Ein weiteres Beispiel sind die für den reibungslosen Betrieb des Internets notwendigen Border Gateway Protocol (BGP)-Router. Der Ausfall bestimmter BGP-Router kann unter Umständen viele Internetnutzer nicht erreichbar machen. Dies können Kunden eines ISP sein oder ein ganzes Volk⁴.

Die Idee der Einführung eines Internet-Kennzahlensystems ist, das komplexe Gebilde des Internets transparenter zu machen und dessen Befinden, Veränderungen und zukünftiges Potential auszudrücken. Die Internetwirtschaft erhält ein gemeinsames Internet-Kennzahlensystem, mit dem der aktuelle Zustand des Internets bezüglich verschiedener Maßstäbe (siehe Abschnitt „Anforderungen an und Mehrwerte eines Internet-Kennzahlensystems“) dargestellt werden kann.

2.2 Kaum statistische Kennzahlen

Es gibt kaum statistische Kennzahlen für die kritische Infrastruktur Internet. Zwar existieren lokale Datensammlungen, jedoch keine in einem umfassenderen Kontext.

Beispielsweise können große Webseitenbetreiber oder Dienste zur Webtraffic-Analyse (wie etwa „Alexa Internet“⁵) Aussagen über die Verbreitung der von den Webseitenbesuchern verwendeten Betriebssystemen, Webbrowser, Software und dergleichen treffen. E-Mail-Anbieter oder Blacklisten-Betreiber können Statistiken zum aktuellen Spamaufkommen liefern. Darüber hinaus stellen Projekte wie „Route Views“⁶, BGPmon⁷ oder „RIPE Atlas“⁸ Informationen bezüglich der Verbindung Autonomer Systeme bzw. der Verfügbarkeit von Internetdiensten dar.

Die verschiedenen Informationen zu Teilaspekten des Internets bieten jedoch aufgrund des fehlenden globalen bzw. umfassenden Charakters nur eine begrenzte Aussage. Viele Erkenntnisse können erst generiert werden, wenn verschiedene Daten miteinander verknüpft werden. Beispielsweise hat eine Meldung über eine schwerwiegende Sicherheitslücke in einem Webbrowser mehr Relevanz, wenn die Software auch tatsächlich von vielen Anwendern genutzt wird. Wenn es diese globale oder zumindest „höhere“ Sicht gäbe, dann könnte der aktuelle Zustand des Internets gemessen, die Entwicklung des Internets besser eingeschätzt und somit fundiertere Entscheidungen für die Zukunft getroffen werden.

3 Anforderungen und Mehrwerte

Es wird ein Internet-Kennzahlensystem benötigt, das eine möglichst umfassende Sicht auf das Internet bereitstellen kann. Es sollen viele verschiedene Teilaspekte des Internets behandelt und zudem die einzelnen Erkenntnisse miteinander verknüpft werden können.

Im Folgenden werden drei grundsätzliche Anforderungen an ein Kennzahlensystem für das Internet definiert: Datenerfassung, Datenverarbeitung und Datenvisualisierung.

³Vgl. unter anderem die Routing-Tabelle des Projekts „Route Views“ unter routeviews.org.

⁴Als Beispiel kann unter anderem Libyen dienen, das im Frühjahr 2011 für einige Zeit komplett vom restlichen Internet getrennt war. Vgl. unter anderem <http://www.heise.de/netze/meldung/Internet-Abschaltung-Libyen-hat-von-Aegypten-gelernt-1206016.html>

⁵<http://www.alexa.com/>

⁶<http://www.routeviews.org/>

⁷<http://bgpmon.net/>

⁸<http://atlas.ripe.net/>

3.1 Datenerfassung

Ein Internet-Kennzahlensystem muss bei der eigentlichen Datenerfassung in der Lage sein, **unterschiedlichste Datenquellen** auslesen zu können. Auch gilt es in Bezug auf die Aussagekraft der erhobenen Daten möglichst allgemeingültige Daten zu erheben. Aktive Messungen müssen beispielsweise auch geografische Hintergründe oder unterschiedliche Technologien berücksichtigen. Verfügbarkeits- und Qualitätsmessungen müssen breit gefächert von verschiedenen Standorten und bestenfalls auch über verschiedene Zugangstechnologien stattfinden.

Die Datenquellen besitzen zudem unterschiedliche Anforderungen bezüglich der **Aktualität der Kennzahlen**, welche durch das Internet-Kennzahlensystem erfüllt werden müssen. Wir teilen die Anforderungen in drei unterschiedliche Zeitbereiche ein. Der kürzeste Zeitbereich bewegt sich im Minuten- bis Stundenbereich. Kennzahlen mit dieser Anforderung können schon bei kurzfristigen Änderungen ein erhebliches Ausmaß an Auswirkung auf das Internet-Kennzahlensystem haben. Insbesondere Messdaten zur Verfügbarkeit von Diensten im Internet spielen hierbei eine große Rolle. Der mittlere Zeitbereich umfasst Tage bis Monate. Hier werden Kennzahlen einsortiert, die durch Änderungen mittelfristig Auswirkungen auf das Internet-Kennzahlensystem zeigen. Dies können unter anderem Routing-Daten des Boarder Gateway-Protokolls sein. So ließe sich beispielsweise der Grad der „Vermaschung“ der verschiedenen Autonomen Systeme beobachten. Beim dritten Zeitbereich handelt es sich um langfristige Daten, welche sich nur selten oder über längere Zeiträume ändern. Eine Erhebung ist hier im Bereich von Quartalen oder Jahren gedacht. Interessant sind hierbei Daten zum Ausbau der Infrastruktur des Internets, wie etwa die durchschnittliche Bandbreite der Endkundenanschlüsse.

Basierend auf den verschiedenen Zeitbereichen ergeben sich **verschiedene Messmethoden**. Kurz- und mittelfristige Datenaktualisierungen sind praktisch nur durch automatisierte Verfahren realisierbar. Ein Internet-Kennzahlensystem muss daher in der Lage sein, selbstständig Daten zu erheben. Zusätzlich muss in der Richtung des Datenflusses unterschieden werden. Daten können entweder von externen Sensoren oder Systemen an das Internet-Kennzahlensystem gesendet werden, oder selbstständig vom Kennzahlensystem angefordert werden. Der Abschnitt „Erhebung von Internet-Kennzahlen“ geht auf diesen Sachverhalt näher ein.

Um verschiedene Kennzahlen vergleichen und die gesammelten Informationen interpretieren zu können, muss ein bestimmter **Geltungsbereich** („Scope“) definiert werden. Dies kann das globale Internet oder ein Teilbereich wie beispielsweise ein „Internet Deutschland“ sein (siehe Abschnitt „Internet-Kennzahlen für Deutschland“).

Im Zusammenhang mit dem Geltungsbereich des Internet-Kennzahlensystems steht die Definition von **Maßstäben**. Es wird bestimmt, welche Bereiche bzw. Schichten des Internets gemessen und behandelt werden. Definierte Maßstäbe gliedern grob ein Internet-Kennzahlensystem (siehe Abschnitt „Internet-Kennzahlen für Deutschland“).

3.2 Datenverarbeitung

Ein Internet-Kennzahlensystem benötigt ein **Analyse- und Bewertungsmodul**, mit dem die gesammelten Informationen verarbeitet werden können. Verschiedene Algorithmen zur Datenanalyse, insbesondere solche aus dem Bereich des Data Mining, sind notwendig. Zudem werden Methoden benötigt, um den „Normalzustand“ des gemessenen Teil des Internets beschreiben zu können. Aufbauend darauf sind Techniken zur Anomalieerkennung nötig, um Veränderungen in den Kennzahlen automatisch detektieren zu können. Nicht nur der Ist-

Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen 5
Zustand, sondern auch die Entwicklung der Kennzahlen soll abgeschätzt und bewertet werden können.

Zusätzlich muss ein Internet-Kennzahlensystem in der Lage sein **unterschiedliche Zusammenhänge** zwischen den Kennzahlen hervorheben zu können. Dies können beispielsweise logische Zusammenhänge („Bedrohungspotential = Bedrohung / Nutzungsgrad“, „Verschlüsselt = Gesamt - Unverschlüsselt“), empirische Zusammenhänge („höhere Nutzung des Firefox-Browsers verursacht mehr AES/SHA1 bei SSL“) oder hierarchische Zusammenhänge („mehr Bot-Netz-Aktivität führt zu mehr DDoS-Attacken führt zu mehr SYN-Anfragen“) sein.

Schließlich müssen die gesammelten Informationen im **Backend** des Internet-Kennzahlensystems sachgerecht hinterlegt werden. Dabei können unter Umständen sehr viele Informationen anfallen, weswegen ein Mechanismus nötig ist, der neben ermittelten Kennzahlen auch aussagekräftige Detaildaten datensparsam speichert. Dies ermöglicht eine nachträgliche Betrachtung von Zusammenhängen, die im Vorfeld – also zum Zeitpunkt der Datenerfassung – noch nicht bedacht wurden oder bekannt waren.

3.3 Datenvisualisierung

Ein Internet-Kennzahlensystem muss in der Lage sein, die gesammelten und zusammengeführten Daten gezielt visualisieren zu können. Es werden verschiedene Instrumente benötigt, um **unterschiedliche Sichten** auf die Kennzahlen zu ermöglichen. Expertenwerkzeuge zur Detailanalyse ermöglichen dem Anwender bestimmte Kennzahlen anzuzeigen (beispielsweise als Zeitreihe oder Kreisdiagramm) und vergleichen zu können. Es muss leicht möglich sein, Vermutungen über Zusammenhänge (siehe Abschnitt „Datenverarbeitung“) verschiedener Kennzahlen überprüfen zu können. Ein Informationsportal bietet einen Gesamtüberblick, ohne in die Tiefe zu gehen. Echtzeitrelevante Informationen können auf das Wesentliche reduziert werden und zum Beispiel als Barometer zugänglich gemacht werden. Schließlich sollte ein Internet-Kennzahlensystem Reporting-Funktionen zur Verfügung stellen. Ein Anwender sollte in der Lage sein, umfangreiche Reports über beliebige Zeitbereiche und Datensätze erstellen zu können, um beispielsweise eine schnelle Untersuchung oder Rekapitulation vergangener Entwicklungen durchführen zu können. Sämtliche Möglichkeiten zur Informationsanzeige sollten für unterschiedliche Endgeräte ausgelegt sein, darunter fallen Implementierungen als Webanwendung oder „App“.

4 Erhebung von Internet-Kennzahlen

Entsprechend der aufgezeigten Anforderungen an ein Internet-Kennzahlensystem ergibt sich die Notwendigkeit unterschiedlichste Datenquellen zu erschließen und deren Daten in das Kennzahlensystem einzuspeisen. Dabei kann zwischen Datenquellen unterschieden werden, deren Daten eingeliefert oder abgerufen werden.

Bei **Einlieferungsdatenquellen** handelt es sich um Datenquellen, die selbstständig Daten in das Kennzahlensystem einspeisen. Dies kann durch Messsensoren geschehen, welche ihre Messwerte über einen Mittler an eine entsprechende Datenquellenimplementierung des Kennzahlensystems übertragen. Es kann sich dabei aber auch um jegliche selbstlaufende Software handeln, welche Daten an das Kennzahlensystem senden möchte. Dies schließt eine Eingabemaske für den manuellen Eintrag von Daten explizit mit ein.

Abruflisten dienen dazu, Daten aus (öffentlich) verfügbaren Datenspeichern abzurufen und in das Kennzahlensystem zu übernehmen. Dazu werden die Datenquellen periodisch durch

6 Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen
das Kennzahlensystem aufgerufen. Hierbei ist eine spezifische Implementierung für jede abzurufende Datenquelle notwendig, was als eigenständiges, durch das Kennzahlensystem aufzurufendes Programm oder als Teil des Kennzahlensystems geschehen kann.

Abbildung 1 zeigt eine mögliche Architektur eines Internet-Kennzahlensystems unter Berücksichtigung der oben erwähnten Datenquellen. Externe Sensoren erfassen aktiv Informationen über den aktuellen Zustand des Internets. Mittels eines Transfersystems werden die gemessenen Informationen in Form eines festgelegten Transferformates über eine passende Einlieferungsdatenquelle an das Kennzahlensystem übermittelt. Für externe Datenbanken und APIs werden Abrufdatenquellen verwendet, welche als spezialisierte Implementierung unter Verwendung spezifischer Konnektoren für den Abruf der Daten genutzt werden können. Weitere Datenquellen lassen sich unter Angabe der zu verwendenden Implementierung in das Backend des Internet-Kennzahlensystems einpflegen.

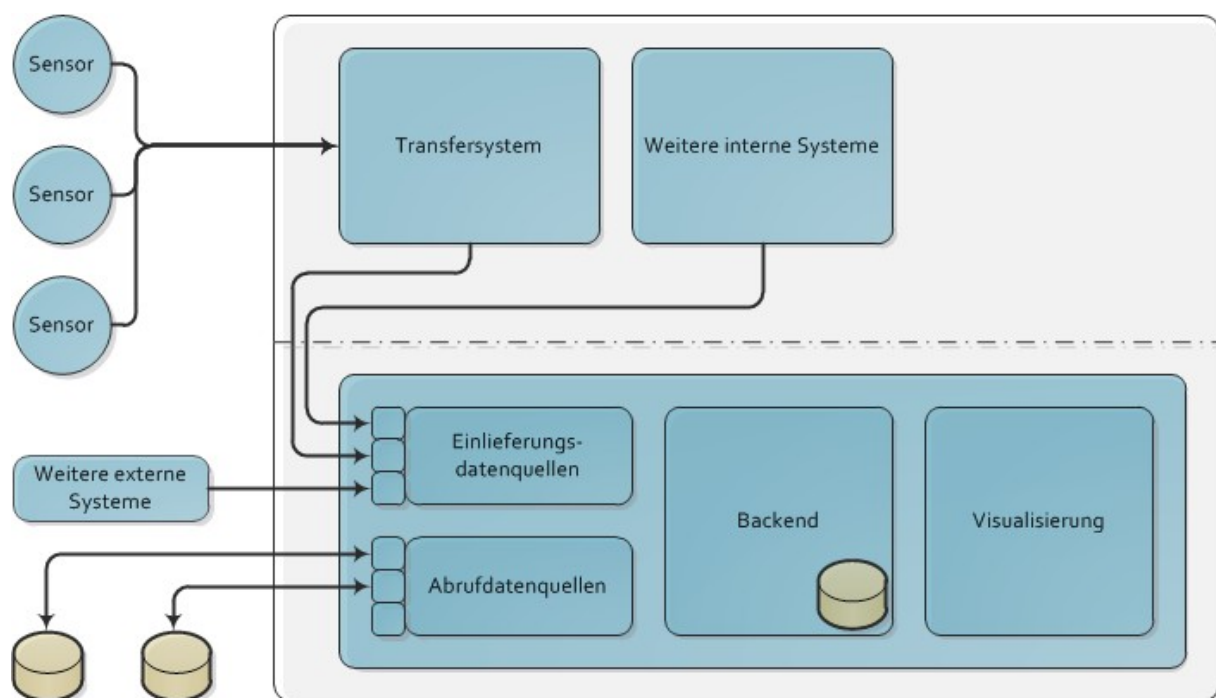


Abbildung 1: Mögliche Architektur eines Internet-Kennzahlensystems.

4.1 Datenquellen

Wie bereits erläutert, kann grundsätzlich zwischen Einlieferungs- und Abrufdatenquellen unterschieden werden. Ein Sondertyp unter den Einlieferungsdatenquellen stellt die manuelle Einlieferung dar. Dadurch lässt sich die Gesamtheit der Datenquellen noch einmal in die Kategorien automatisierbare und manuelle Datenquellen unterteilen.

4.1.1 Automatisierbare Datenquellen

Bei den Datenquellen wird zwischen automatisierbaren und manuellen Datenquellen unterschieden. Zu solchen, die automatisiert eingepflegt werden können, gehören:

Verfügbarkeitskennzahlen

Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen 7

Dazu zählen Daten, die über die Verfügbarkeit wichtiger Dienste des Internets gesammelt werden. Beispielsweise werden Webseiten und Dienste hinsichtlich ihrer Verfügbarkeit und Benutzbarkeit für den Endnutzer (Quality of Service, QoS) überprüft. Die Interpretation der QoS-Datenbasis im Sinne einer etwaigen Quality of Experience (QoE) stellt ebenfalls einen interessanten Aspekt dar, sofern hier eine Abschätzung möglich ist. Ein Beispiel für eine Datenquelle, die Verfügbarkeitskennzahlen erheben kann, ist das Projekt „RIPE Atlas“ oder auch das Internet-Verfügbarkeitssystem [Oste06].

Statistiken zu Protokollen und Technologien

Auf der Basis von anonym und datenschutzkonform gesammelten Messdaten von zentralen Internet-Infrastrukturpunkten ist es möglich, Nutzungsstatistiken von Protokollelementen und Technologien zu generieren. Die dabei interessanten Aspekte sind beispielsweise die Verteilungen der verwendeten Webbrowser, Betriebssysteme oder das Verhältnis von IPv4 zu IPv6. Beispiele für solche Datenquellen sind das Projekt „HTTP Archive“⁹ oder das Internet-Analyse-System [Proe05].

Sicht auf die Verflechtung des Internets

Durch das Abrufen frei verfügbarer BGP-Routingdaten wird ein Überblick der Autonomen Systeme des Internets ermöglicht. Dabei ist für die Generierung von Infrastrukturkennzahlen vor allem die Verknüpfung der AS untereinander von Interesse. Außerdem bieten die Basisdaten die Möglichkeit einer gezielten Recherche durch eine komfortable Sicht auf die Routing-Datenbasis. Eine solche Datenquelle ist beispielsweise die Datenbank des „Route View“-Projekts oder der aiconViewer [Dier06].

Nutzungsstatistiken durch APIs

Unterschiedliche Dienstanbieter im Internet verfügen über eine massive Datenbasis bezüglich Teilaspekte des Internets und stellen Teile daraus frei im Internet zum Abruf über APIs zur Verfügung. Diese Daten bieten sich ebenfalls als Teil eines Internet-Kennzahlensystems an. Ein Beispiel für eine solche Datenbasis ist die Google Safe Browsing API¹⁰, welche es ermöglicht, die Verbreitung von Phishing- oder Malware-Seiten innerhalb eines Autonomen Systems festzustellen.

4.1.2 Manuelle Datenquellen

Neben den automatisierbaren Datensätzen gibt es durchaus wichtige Informationen, die manuell in das Internet-Kennzahlensystem eingepflegt werden müssen. Dazu gehören:

Infrastrukturdaten

Statistiken oder andere Reporte können genutzt werden, um beispielsweise die Verbreitung von DSL-Anschlüssen, verschiedener Bandbreiten, mobiler Endgeräte oder auch Internetzugangstechnologien zu erheben. Die Daten können hierbei über die manuelle Oberfläche des Internet-Kennzahlensystems eingetragen werden. Ein Beispiel für eine solche Datenquelle sind die Statistiken der deutschen Bundesnetzagentur.

Informationen zum Bedrohungspotential

Öffentlich verfügbare Schwachstellen-Datenbanken von Sicherheitsinitiativen oder Sicherheitsbarometern können genutzt werden, um beispielsweise die Verbreitung von Sicherheitslücken, Schäden durch Internetkriminalität oder Spamaufkommen zu analysieren. Die Daten können hierbei über die manuelle Oberfläche eines Internet-Kennzahlensystems

⁹<http://httparchive.org/>

¹⁰<http://code.google.com/intl/de-DE/apis/safebrowsing/>

8 Ein Internet-Kennzahlensystem für Deutschland: Anforderungen und technische Maßnahmen eingetragen werden. Ein Beispiel für eine solche Datenquelle ist die Schwachstellen-Datenbank des National Institute of Standards and Technology (NIST)¹¹.

4.2 Weitere technische Maßnahmen

Auch wenn das Kennzahlensystem als solches nur Kennzahlen speichert und bereitstellt, kann es sinnvoll sein in den zuliefernden Datenquellen detailliertere Daten vorzuhalten. Dies soll dazu dienen, eine genauere Auswertung von im Kennzahlensystem beobachteten Effekten zu ermöglichen. Dazu muss die Möglichkeit einer Neubetrachtung der Basisdaten im Zeitverlauf ermöglicht werden. Bei diesen detaillierteren Daten kann es sich nicht um die Rohdaten handeln, welche für die Auswertung genutzt wurden, da dies im Sinne des Speicherplatzes und des Datenschutzes zu Problemen führen kann. Es gilt also sparsam mit Daten umzugehen, ohne an Aussagekraft zu verlieren, und in keinem Fall datenschutzrechtlich bedenkliche Daten vorzuhalten.

Zu den mächtigsten Analyse- und Statistikwerkzeugen für die Art an Informationen, die für ein Kennzahlensystem von Relevanz sind, gehören solche, die mittels Deep-Packet-Inspection große Mengen an Internetverkehr „sehen“ und auswerten können. Dies kann jedoch schnell zu Problemen mit dem Datenschutz führen. So ist auch das Aufzeichnen ganzer Datenströme, was zum maximalen Informationsgewinn führen würde, unzulässig und impraktikabel. Um datenschutzkonform und performant an Daten zur aktuellen Internetnutzung zu gelangen ist es notwendig, entsprechend dem Datensparsamkeitsprinzip vorzugehen, und außerdem keine datenschutzrechtlich bedenklichen Informationen zu betrachten oder gar zu persistieren.

Für die statistische Erhebung von Internetverkehr nach den erwähnten Kriterien der Performanz und Datenschutzunbedenklichkeit, wie sie für ein Kennzahlensystem sinnvoll ist, bietet sich beispielsweise die Verwendung von sFlow-Datenströmen in Verbindung mit einer geeigneten datenschutzkonformen Auswertungssoftware an. Dies bietet den Vorteil, dass die Auswertungssoftware durch sFlow mit einer statistisch aussagekräftigen Menge an Paketen versorgt werden kann, aber gleichzeitig an Knotenpunkten mit großen Datenmengen kein Performanceproblem entsteht. Dies wird dadurch gewährleistet, dass mittels sFlow zum einen von verschiedensten Netzwerkkomponenten Pakete an die Auswertungssoftware geschickt werden können. Zum anderen kann durch die Samplerate, mit der sFlow die Anzahl der an die Auswertungssoftware geschickten Pakete limitiert, eine Überlastung der Auswertungssoftware verhindert werden. Außerdem wird durch die Paketlängenlimitierung von sFlow das Datenschutzproblem dahingehend entschärft, dass nur die ersten 128 Byte eines Paketes zu der Auswertungssoftware übertragen werden, und somit ein Großteil der sensiblen Payload-Daten gar nicht erst in der Auswertungssoftware auftreten.

Die Herausforderung besteht darin, datensparsam und datenschutzkonform eine hohe Aussagekraft für die Betrachtung im Zeitverlauf zur Verfügung zu stellen, um später auf dieser Datenbasis mittels eines Expertensystems auch neue Aussagen treffen zu können.

¹¹<http://nvd.nist.gov/>

5 Internet-Kennzahlen für Deutschland

Die in dieser Arbeit beschriebene Notwendigkeit für ein Internet-Kennzahlensystem ist auch konkret für den Staat Deutschland gegeben. Es wird ein System benötigt, mit dessen Hilfe Internet-Kennzahlen gesammelt und aufbereitet werden, woraus Handlungsoptionen für die Politik, partizipierende Unternehmen und Bürger generiert werden können. Mit Hilfe eines etablierten Internet-Kennzahlensystems kann anschließend evaluiert werden, ob durch umgesetzte Maßnahmen die gewünschte Wirkung erzielt wurde.

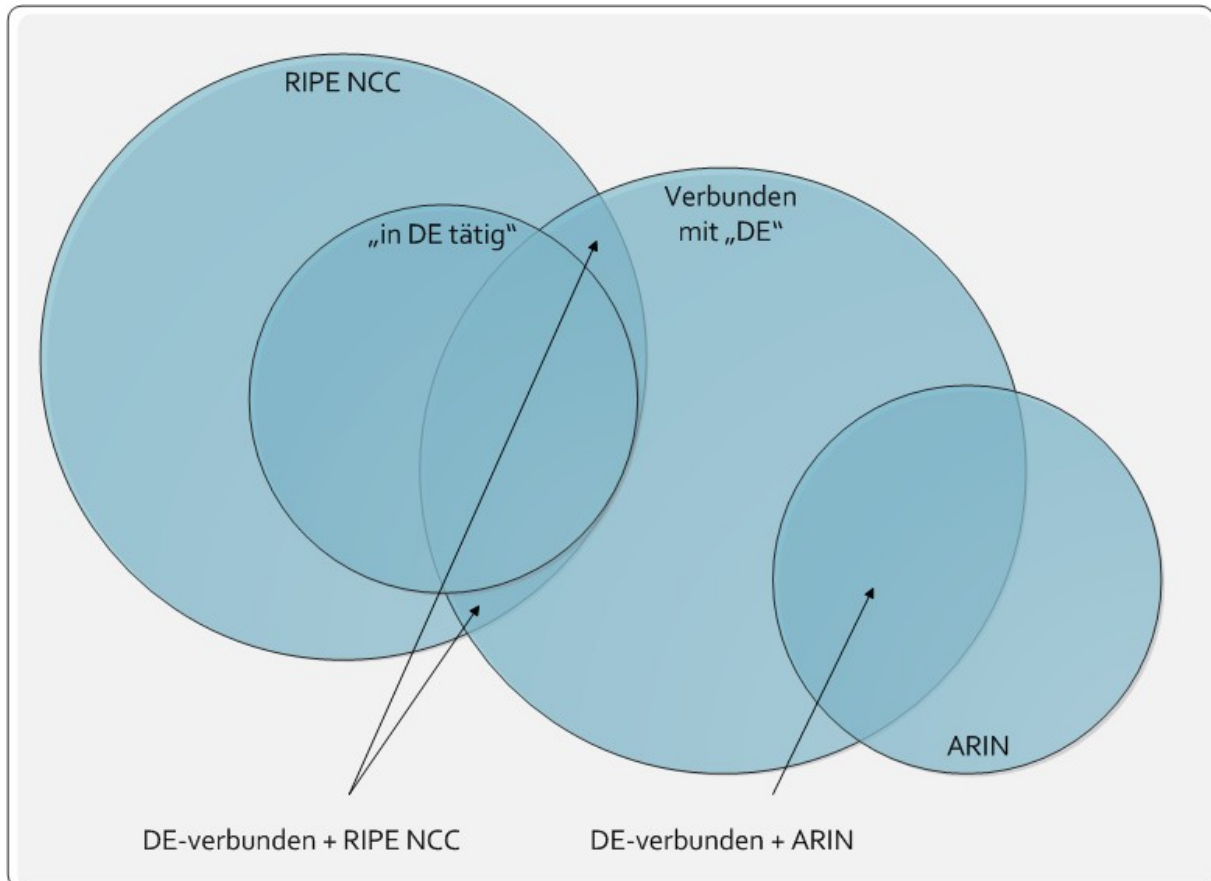


Abbildung 2: Mögliche Definition eines „Internet Deutschland“ über Autonome Systeme.

Eine der wichtigsten Voraussetzungen für ein Kennzahlensystem für Deutschland ist die Definition eines **Geltungsbereiches** (siehe Abschnitt „Anforderungen an und Mehrwerte eines Internet-Kennzahlensystems“). Eine Möglichkeit, ein „Internet Deutschland“ zu definieren, bezieht sich auf die Autonomen Systeme (vgl. Abbildung 2):

Von den derzeit etwa 38.000 aktiven Autonomen Systemen im Internet sind nur solche relevant, bei denen der Betreiber in Deutschland tätig ist. Dies kann beispielsweise über die geographische Verwendung der vergebenen IP-Adressen (zum Beispiel über eine GeoIP-Datenbank), oder über die Länderzuordnung der RIPE-Datenbank¹² abgeglichen werden. Darüber hinaus werden solche Autonomen Systeme zum „Internet Deutschland“ gezählt, die

¹²Vgl. <ftp://ftp.ripe.net/pub/stats/ripenncc/delegated-ripenncc-latest>

direkt mit den zuvor genannten AS verbunden sind und sich zusätzlich in Europa oder Nordamerika befinden. Dies entspricht der Registrierung bei dem RIPE NCC oder der ARIN.

Zudem kann ein Internet-Deutschland in vier **Akteure bzw. Aspekte** eingeteilt werden. Es handelt sich um die Infrastruktur (AS mit BGP-Routern, Verbindungen, ...), Dienste (E-Mail, Web, VoIP, ...), Teilnehmer (PCs, Notebooks, Smartphones, Autos, Kühlschränke, ...) und Bedrohungen (Malware, Botnetze, (Distributed) Denial of Service-Angriffe, ...). Aus den vier Aspekten werden die **Maßstäbe** abgeleitet, die zu den Anforderungen an ein Internet-Kennzahlensystems gehören:

- **Leistungsfähigkeit:** Repräsentiert den Aspekt Infrastruktur im Kennzahlensystem und umfasst Parameter, die Aussagen über die Kapazität und Abhängigkeiten des Internets treffen. Zu solchen Parametern zählen unter anderem die durchschnittliche Hop-Anzahl, Bandbreite und Paketverlustrate. Ferner ist die Information relevant, welcher Anteil der Autonomen Systeme ohne die Nutzung von Transit erreicht werden kann.
- **Verfügbarkeit:** Repräsentiert den Aspekt der Dienste im Kennzahlensystem und umfasst Aussagen über die Verfügbarkeit von Diensten im Internet aus dem Blickwinkel des Endanwenders.
- **Einschätzung der Nutzung:** Repräsentiert den Aspekt Teilnehmer im Kennzahlensystem und umfasst Parameter, die Aussagen über die Nutzung des Internets und die verwendeten Technologien treffen. Dazu gehören beispielsweise die Verbreitungsgrade von Betriebssystemen, Browser-Software und Zugangstechnologien, wie DSL oder UMTS, sowie die Aufschlüsselung des Gesamtdatenaufkommens.
- **Bedrohungspotential:** Repräsentiert den Aspekt Bedrohungen im Kennzahlensystem und umfasst Parameter, die Aussagen über die Gefährdung des Internets treffen. Als Gefährdungsindikator gilt beispielsweise die Anzahl neuer Virensignaturen pro gemessenen Zeitraum, die Anzahl infizierter Webseiten oder die gemessenen Datenraten bei DDoS-Angriffen.

Ein soeben beschriebenes Internet-Kennzahlensystem wird viele Kennzahlen sammeln und viele Analysen und Interpretationen durchführen, aber auch eine gute Basis für weitere Arbeiten und Betrachtungen durch Dritte bieten. Eine Mitarbeit ist somit seitens der benötigten Kennzahlen möglich sowie durch die Nutzung des Systems durch Dritte.

6 Fazit und Ausblick

Ein Internet-Kennzahlensystem, welches die in dieser Arbeit beschriebenen Anforderungen erfüllt, ist enorm wichtig für den fortwährenden Betrieb des Internets. Insbesondere kann eine Fokussierung auf einen bestimmten Teilbereich des Internets, zum Beispiel auf ein „Internet Deutschland“, konkrete Mehrwerte für Nutzer und Teilnehmer bieten. So kann mittels eines Internet-Kennzahlensystems der aktuelle Zustand des Internets gemessen, Normalzustände definiert und so schnell Abweichungen und Anomalien erkannt werden. Zusammenhänge von Kennzahlen, die nicht ohne weiteres ersichtlich sind, können mithilfe verschiedener Werkzeuge konkretisiert und analysiert werden. Zudem erlaubt die dabei stattfindende Datenerhebung eine retrospektive Betrachtung der Entwicklung des Internets aus neuen Blickwinkeln.

Problematisch bei der Realisierung eines Internet-Kennzahlensystems ist weniger die zugrunde liegende Technik, sondern vielmehr die benötigten Partner. Neben den vielen frei verfügbaren Informationen sind auch zahlreiche weitere Kennzahlen von Unternehmen, Behörden oder Initiativen von Interesse und auch notwendig. Diese sind jedoch in vielen Fällen nicht öffentlich

zugänglich und müssen durch individuelle Absprachen in das Internet-Kennzahlensystem integriert werden.

Danksagung

Diese Arbeit ist Teil des Projektes "Deutscher Internet-Index (DIX)", welches vom Bundesministerium für Wirtschaft und Technologie (BMWi) finanziert wird. Der Inhalt dieser Veröffentlichung steht in alleiniger Verantwortung der Autoren und widerspiegelt in keiner Weise die Meinung des BMWi.

Literatur

- [BMI09] Bundesministerium des Innern: Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie). Berlin, 12. Juni 2009.
- [Dier06] Stefan Dierichs: Eine strukturelle Analyse des Internet - Zusammenhänge der Internetkommunikation mit der besonderen Betrachtung des Standortes Deutschland. Diplom-Arbeit. Fachhochschule Gelsenkirchen, 2006.
- [Kuet03] Martin Kütz: Kennzahlen in der IT – Werkzeuge für das Controlling und Management. dpunkt Verlag, 2003.
- [Oste06] Thomas Ostermann: Internet-Verfügbarkeitssystem. Diplom-Arbeit. Fachhochschule Gelsenkirchen, 2006.
- [Proe05] Marcus Proest: Entwicklung einer Sonde für ein Internet-Analyse-System. Diplom-Arbeit. Fachhochschule Gelsenkirchen, 2005.
- [Wott10] Tobias Wottawa: Ein Model zur Visualisierung des Internet Bandbreitenbedarfs in Deutschland. Bachelor-Arbeit. Fachhochschule Gelsenkirchen, 2010.