Mehrwert eines Internet-Kennzahlensystems für Deutschland

Deutscher Internet-Index

Das Internet ist eine kritische Infrastruktur, deren fortwährender Betrieb besonders schützenswert ist. Problematisch bei der Betrachtung des Zustandes des Internets sind zwei Dinge: Zum einen gibt es viele Abhängigkeiten im Kontext des Internets, zum anderen aber nur wenige Kennzahlen, mit denen umfassende Aussagen getätigt werden können. Dieser Artikel beschreibt ein Kennzahlensystem, bei dem das Steuerungsobjekt das Internet als solches ist. Das komplexe Gebilde Internet soll transparenter gemacht sowie dessen Befinden, Veränderungen und zukünftiges Potential ausgedrückt werden. Neben der grundsätzlichen Idee, die der Konzeption und Realisierung eines Internet-Kennzahlensystems zu Grunde liegt, wird auch der Mehrwert für die verschiedenen Adressaten erläutert.

Das heutige Internet besteht aus einer Vielzahl von Netzwerken, sogenannten „autonomen Systemen" (AS), die von Internet Service Providern, großen Unternehmen oder Hochschulen betrieben werden. Zurzeit gibt es mehr als 38.000 AS, die mit mehr als 70.000 Verbindungen das Internet bilden. Die autonomen Systeme, die sich in Größe und räumlicher Ausdehnung immens unterscheiden können, werden unterschiedlich und vollkommen unabhängig voneinander verwaltet. Das bedeutet, dass die Betreiber beispielsweise unterschiedliche Strategien haben, wie sie mit Hilfe von Routing-Protokollen die Kommunikation der IP-Pakete in ihren Netzen organisieren. Darüber hinaus zählt das Internet als Teil der Informations- und Telekommunikationstechnologie nicht nur innerhalb Deutschlands zu den kritischen Infrastrukturen. Die Beeinträchtigung oder der Ausfall von Teilen des Internets kann enorme Auswirkungen haben.

Motivation für die Erhebung von Internet-Kennzahlen

Grundsätzlich lässt sich die Notwendigkeit von Internet-Kennzahlen in zwei Kategorien einteilen. Zum einen besitzt das Internet viele unterschiedliche Abhängigkeiten. Diese können auf der Ebene der Technik, der Dienste oder der Verwaltung existieren. Die Idee der Verwendung von Internet-Kennzahlen ist, das komplexe Gebilde des Internets transparenter zu machen und dessen Befinden, Veränderungen und zukünftiges Potential auszudrücken.

Zum anderen gibt es kaum statistische Kennzahlen, die die kritische Infrastruktur Internet umfassend beschreiben. Zwar existieren lokale Datensammlungen, jedoch sind diese oft nicht-öffentlich oder sie haben keinen Bezug zueinander. Die verschiedenen Informationen zu Teilaspekten des Internets bieten aufgrund des fehlenden globalen beziehungsweise umfassenden Charakters nur eine begrenzte Aussage. Viele Erkenntnisse können erst generiert werden, wenn verschiedene Daten miteinander verknüpft werden.

Grundsätzliche Idee eines Systems zur Verarbeitung von Internet-Kennzahlen

Es wird ein System benötigt, das eine möglichst umfassende Sicht auf das Internet bietet. Es sollen viele verschiedene Teilaspekte des Internets behandelt und zudem die einzelnen Erkenntnisse miteinander verknüpft werden. Um die Komplexität des Internets analysieren zu können, ist der Einsatz eines Kennzahlensystems sinnvoll, wie es bereits in der Betriebswirtschaft

breite Anwendung findet. Nur ist eben der zu untersuchende Sachverhalt kein Unternehmen, sondern das Internet als solches. Es werden Kennzahlen gesammelt, die sich auf das Internet beziehen oder die durch das Internet generiert werden.

Der von einem Internet-Kennzahlensystem gemessene Teil des Internets kann in unterschiedliche Aspekte beziehungsweise Akteure eingeteilt werden, welche alle an dem Internet beteiligten Elemente beinhalten. Aus den Aspekten werden diejenigen Maßstäbe abgeleitet, über die ein Internet-Kennzahlensystem Aufschluss geben muss:

Der Maßstab Leistungsfähigkeit repräsentiert den Aspekt Infrastruktur (AS mit BGP-Router, Verbindungen, …). Er 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 Transitverbindungen erreicht werden kann.

Der Maßstab Verfügbarkeit repräsentiert den Aspekt Dienste (E-Mail, Web, VoIP, …). Er umfasst Aussagen über die Verfügbarkeit von Diensten im Internet aus dem Blickwinkel des Anwenders.

Der Maßstab „Einschätzung der Nutzung" repräsentiert den Aspekt Teilnehmer (PCs, Notebooks, Smartphones, Autos, Kühlschränke, …). Er umfasst Parameter, die Aussagen über die Nutzung des Internets und der verwendeten Technologien treffen. Dazu gehören beispielsweise die Verbreitungsgrade von Betriebssystemen, Browser-Software und Zugangstechnologien wie DSL oder UMTS, sowie die Aufschlüsselung des Gesamtdatenaufkommens.

Der Maßstab Bedrohungspotential repräsentiert den Aspekt Bedrohungen (Malware, Bot-Netze, …). Er 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.

Definition „Internet Deutschland"

Eine wichtige Voraussetzung für ein Kennzahlensystem für Deutschland ist die Definition eines Geltungsbereichs. Die Abgrenzung des zu messenden Bereichs des Internets ist aus zweierlei Gründen notwendig. Zum einen müssen die ermittelten Kennzahlen vergleichbar sein. Bestimmte gemessene Zahlen können mit denen anderer Länder abgeglichen oder gegen die Gesamtmenge gestellt werden. Zum anderen müssen die ermittelten Kennzahlen interpretierbar sein. Erst hierdurch können Zusammenhänge zwischen unterschiedlichen Daten erkannt und bewertet werden. Eine Möglichkeit, ein „Internet Deutschland" zu definieren, bezieht sich beispielsweise auf die Autonomen Systeme.

Erstellung und Betrieb eines Internet-Kennzahlensystems

Ein Internet-Kennzahlensystem besteht grundsätzlich aus drei Komponenten: Datenerfassung, Datenverarbeitung und Datenvisualisierung. Bei der eigentlichen Datenerfassung muss ein Internet-Kennzahlensystem in der Lage sein, unterschiedlichste Datenquellen auszulesen. Auch gilt es in Bezug auf die Aussagekraft der erhobenen Daten möglichst allgemeingültige Daten zu erheben. Die Daten können entweder von externen Sensoren oder Systemen an das Internet-Kennzahlensystem gesendet oder selbständig vom Kennzahlensystem angefordert werden.

Innerhalb der Komponente Datenverarbeitung sollte ein Internet-Kennzahlensystem ein Analyse- und Bewertungsmodul besitzen, mit dem die gesammelten Informationen verarbeitet werden. Es werden Methoden benutzt, um den „Normalzustand" des gemessenen Internets zu beschreiben. Aufbauend darauf sind Techniken zur Anomalieerkennung nötig, um Veränderungen in den Kennzahlen automatisch zu detektieren. Nicht nur der Zustand, sondern auch die Entwicklung der Kennzahlen wird gemessen und bewertet. Zusätzlich sollte ein Internet-Kennzahlensystem in der Lage sein, unterschiedliche Zusammenhänge zwischen den Kennzahlen hervorzuheben. 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 Botnetz-Aktivität führt zu mehr DDoS-Attacken führt zu mehr SYN-Anfragen") sein.

Mittels der Datenvisualisierung ist ein Internet-Kennzahlensystem in der Lage, die gesammelten und zusammengeführten Informationen gezielt darzustellen. Es werden verschiedene Instrumente benötigt, um unterschiedliche Sichten auf die Kennzahlen zu ermöglichen. Neben Expertenwerkzeugen zur Detailanalyse sollte auch ein Informationsportal für einen Gesamtüberblick und Reporting-Funktionalitäten zur Verfügung stehen. Es muss leicht möglich sein, Vermutungen über Zusammenhänge verschiedener Kennzahlen überprüfen zu können.

Mehrwert

Der wohl größte Mehrwert eines Kennzahlensystems für das Internet ist die Sammlung und Zusammenführung vieler Informationen, die das Verständnis von der kritischen Infrastruktur Internet fördern. So können beispielsweise Verfügbarkeitskennzahlen ermittelt werden, die Informationen über die Erreichbarkeit von Webseiten und Diensten geben. Statistiken zu Protokollen und Technologien des Internets sind möglich, wenn anonyme Messdaten an zentralen Internet-Infrastrukturpunkten gesammelt und verarbeitet werden. Interessante Aspekte hierbei sind beispielsweise Verteilungen der verwendeten Web-Browser, Betriebssysteme oder das Verhältnis von IPv4 zu IPv6. Es kann eine Sicht auf die Verflechtung des Internets gegeben werden, wenn (frei verfügbare) BGP-Routingdaten abgerufen und verarbeitet werden. Insbesonde-

re die Verknüpfungen der AS untereinander sind interessant. Außerdem verfügen die verschiedensten Dienstanbieter im Internet über eine umfassende Internet-Datenbasis, die sie – zumindest teilweise – wiederum über das Internet frei zur Verfügung stellen. Dadurch sind neue Erkenntnisse mittels Kombination und Analyse der Kennzahlen möglich. Infrastrukturdaten, etwa Informationen über die Verbreitung mobiler Endgeräte, Internetzugangstechnologien und mehr, können durch Statistiken oder Reports verschiedener Institutionen eingebunden werden. Öffentlich verfügbare Datenbanken über Schwachstellen, von Sicherheitsinitiativen oder Sicherheitsbarometern können genutzt werden, um beispielsweise die Verbreitung von Sicherheitslücken, Schäden durch Internetkriminalität oder Spamaufkommen zu erheben.

Mithilfe eines umfassenden Kennzahlensystems steht eine Datenbasis zur Verfügung, die nicht mehr nur lokal existiert, sondern einen umfassenderen Kontext besitzt. Es können Erkenntnisse generiert werden, der aktuelle Zustand des Internets gemessen, die Entwicklung des Internets besser eingeschätzt und somit fundiertere Entscheidungen für die Zukunft getroffen werden. Zusammengefasst können eventuell unbekannte Zusammenhänge besser festgestellt, mit den neu erworbenen Erkenntnissen Prognosen erstellt und auch Entwicklungen erklärt werden.

Ein Internet-Kennzahlensystem wird viele Kennzahlen sammeln und zahlreiche Analysen und Interpretationen durchführen, aber auch eine gute Basis für weitere Arbeiten und Betrachtungen durch Dritte bieten. Eine Mitarbeit ist somit einerseits durch partizipierende Unternehmen und deren zur Verfügung gestellte Kennzahlen möglich. Zum anderen kann das System durch Dritte genutzt werden und dementsprechend selbst als Datenbasis dienen. Anwender stützen sich so nicht nur auf Erfahrung, sondern beziehen sich auch auf konkrete Zahlen von anderen Quellen, deren Interpretation sie eventuell auf ganz eigene Schlüsse bringt.

Ein Internet-Kennzahlensystem bietet eine Fülle von aufbereiteten und komprimierten Informationen, welche von zuständigen Stellen zur Entwicklung geeigneter Handlungsoptionen genutzt werden können. Ziel dieser Empfehlungen sind unterschiedlichste Entscheidungsträger: Die Politik, partizipierende Unternehmen oder Bürger. Außerdem kann evaluiert werden, ob bereits umgesetzte Maßnahmen die gewünschte Wirkung erzielt haben.

- Politik: Gute Informationen sind die Grundlage für eine angemessene Steuerung und richtige Entscheidungen zu regulatorischen und auch rechtlichen Fragen. Aussagen lassen sich nach Möglichkeit durch ein Internet-Kennzahlensystem untermauern oder widerlegen.

- Partizipierende Unternehmen: Gute Informationen lassen Trends und Entwicklungen im Internet-Markt erkennen. Dies verbessert die Grundlage für strategische Entscheidungen. Außerdem können betriebsinterne Daten mit den Internet-Kennzahlen in Beziehung gesetzt und somit angereichert oder gegengeprüft werden.

- Bürger: Entsprechend dargestellte Informationen können das Interesse für Teilgebiete des Internets – beispielsweise der Aspekt Sicherheit – fördern. So kann beispielsweise eine Sensibilisierung der Bürger für das Thema IT-Sicherheit stattfinden, wenn eine starke Zunahme an Phishing-Mails oder infizierten Webseiten dargestellt und entsprechende Warnungen oder Verhaltensweisen ausgegeben werden.

Fazit und Ausblick

Ein Internet-Kennzahlensystem ist ein leistungsfähiges Expertenwerkzeug zur Leistungsüberwachung und Trenderkennung im Internet. Insbesondere kann eine spezialisierte Fokussierung auf einen bestimmten Teilbereich des Internets, zum Beispiel auf ein „Internet Deutschland", Erkenntnisse liefern, die in generalisierten beziehungsweise globalen Systemen nicht ohne weiteres offensichtlich sind. Zudem können der aktuelle Ist-Zustand des Internets gemessen, Normalzustände definiert und schnell Abweichungen beziehungsweise Anomalien erkannt werden.

Die erhobenen Daten erlauben durch ihre Quantifizierung bislang noch unbekannte Zusammenhänge zu entdecken, zu konkretisieren und zu analysieren. Die Art der Kennzahlen, Zusammenhänge und Aussagen ist dabei nicht auf einen bestimmten Aspekt (etwa Wirtschaft oder Technik) beschränkt, sondern hängt maßgeblich davon ab, mit welchen Regeln die Kennzahlen generiert und analysiert werden. Die Zusammensetzung und Auswertung der Kennzahlen bietet dabei ein sehr großes Potential für weitergehende Forschung und Arbeiten.

Fernab davon hängt der Erfolg eines Internet-Kennzahlensystems hauptsächlich von den partizipierenden Unternehmen ab. Wie jedes Kennzahlensystem ist auch ein Internet-Kennzahlensystem nur so gut wie die eingepflegten Daten. Neben den frei verfügbaren Daten ist die kontinuierliche und freiwillige Einbringung von bisher unveröffentlichten Daten absolut wünschenswert und notwendig. Dies optimiert die Aussagekraft des Internet-Kennzahlensystems. ■

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.

Für Abonnenten ist dieser Artikel auch digital auf www.datakontext.com verfügbar

Prof. Dr. (TU NN) Norbert Pohlmann, Informatikprofessor für Verteilte Systeme und Informationssicherheit sowie Leiter des Instituts für Internet-Sicherheit an der Fachhochschule Gelsenkirchen. (www.internet-sicherheit.de)

M.Sc. Sebastian Feld, wissenschaftlicher Mitarbeiter am Institut für Internet-Sicherheit in den Forschungsbereichen Internet-Kennzahlen und Identity Management.