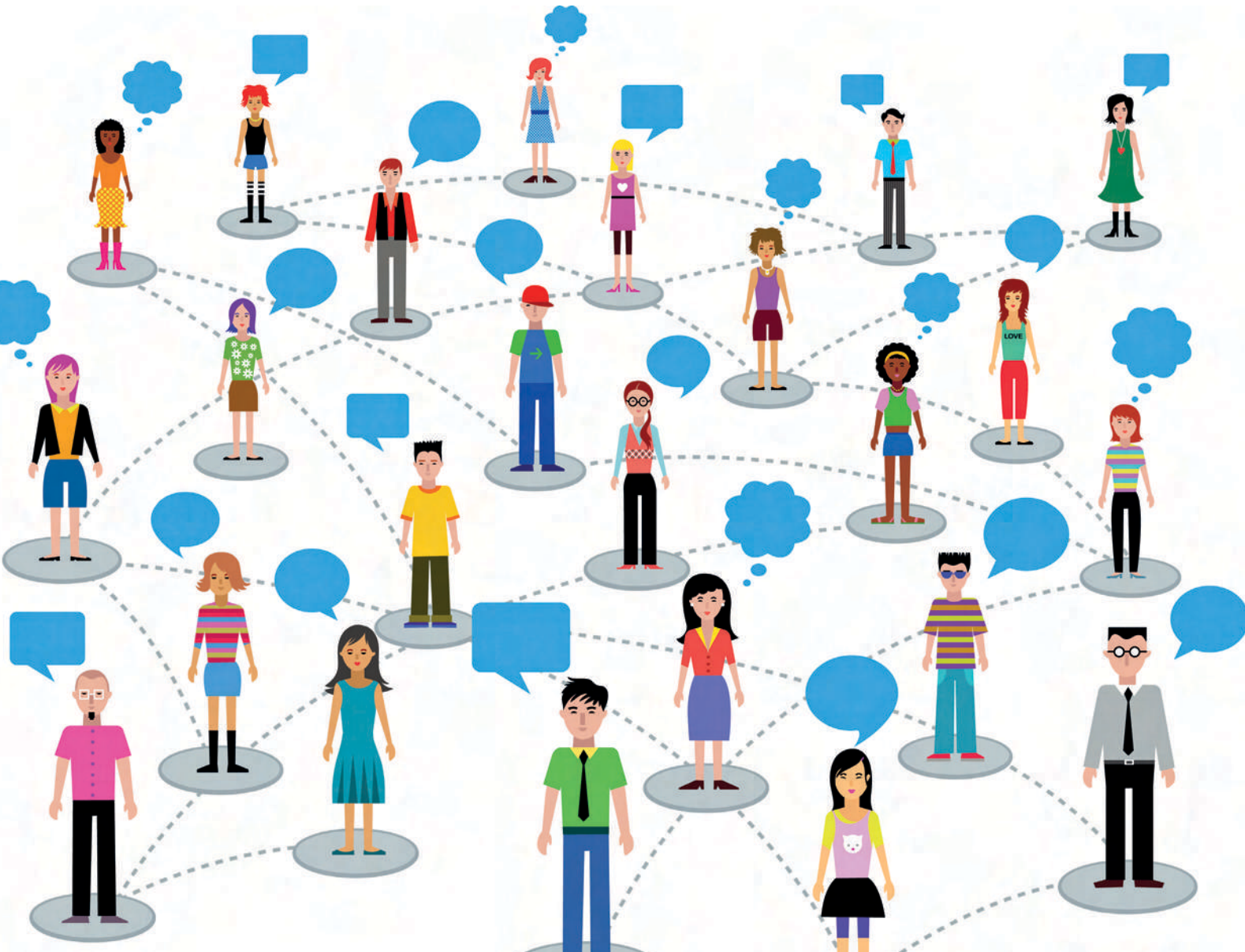
Das Internet als Plattform für „Big Data“ (Teil 1)

Kommunikation als Massenerlebnis

Ohne das Internet funktioniert heute nichts mehr – sowohl im beruflichen Sektor als auch im privaten Bereich. Gerade wenn es darum geht, große Mengen an Daten zum Download anzubieten und autorisiert zu verteilen, reichen klassische Dienste wie das Web oder die E-Mail nicht aus. Es ist daher nicht verwunderlich, dass sich gerade in diesem Bereich in den letzten Jahren sehr viel verändert hat. Diese Kurzserie untersucht die Kommunikationsentwicklung in zwei verschiedenen Modellvarianten. Im vorliegenden ersten Teil geht es um den Austausch großer Datenmengen wie Betriebssysteme, Programme, HD-Videos und Updates. Hierbei kennen sich die beteiligten Kommunikationspartner üblicherweise nicht persönlich und die Daten werden relativ anonym ausgetauscht. Die zweite Variante, Gegenstand des zweiten Teils, basiert auf der persönlichen Kommunikation und ist meist zwischenmenschlich ausgerichtet, beispielsweise in Form des Austauschs privater Fotos und Videos.

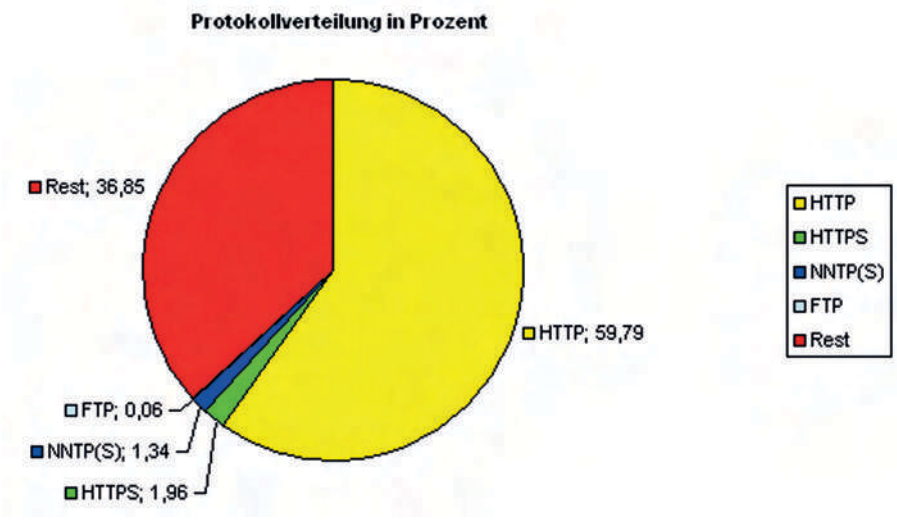
Wenn es darum geht, mit unbekannten Personen und Computern große Datenmengen auszutauschen, funktioniert es in der Regel nicht, Images von Betriebssystemen oder eine Videosammlung per Mail zu versenden. Für diese Form des Datenaustauschs wurden bereits vor etlichen Jahren Protokolle wie das File Transfer Protocol (FTP) entwickelt, die genau dies im Fokus haben. So konnten Unternehmen einen eigenen FTP-Server im Internet platzieren und dort wichtige Daten wie Updates allen interessierten Benutzern zur Verfügung stellen. Technischer Hintergrund ist hierbei, wie bei den meisten Standard-Protokollen, eine Client-Server-Architektur.

Einen ähnlichen Mechanismus verwendet das Network News Transfer Protocol (NNTP). Ursprünglich war dies für eine Art Informations-Pool gedacht: Personen schicken Nachrichten zu einem Server, welcher diese speichert und für andere Mitglieder – meist öffentlich – zur Verfügung stellt. Daraus entstand das so genannte „Usenet“. NNTP kennt bei den Nachrichten einen Betreff sowie einen Textkörper und es können bequem Dateien angehängt werden. Größter Unterschied zur Mail ist jedoch, dass der Server an ihn verschickte



Nachrichten im Prinzip jedem Mitglied einer Gruppe zur Verfügung stellt, also einem ganzen Abonnentenkreis. Heutzutage sind solche Newsgroup-Server und das Usenet in der Bedeutung stark zurückgegangen. NNTP und die verschlüsselte Variante NNTPS werden jedoch vermehrt zum Datenaustausch verwendet, oft auch mit illegalem Inhalt. Hintergrund ist die bequeme Server-zu-Client-Verteilung und die eingebaute Möglichkeit zur Authentifizierung, sodass zum Beispiel nur zahlende Kunden Zugang erhalten. Der Datentransfer unter den Teilnehmern kann zudem bei NNTPS noch mittels SSL/TLS während der Übertragung verschlüsselt werden.

Sollen viele Daten verteilt werden, stellt sich die Client-Server-Architektur oft als eine aufwändige und teure Angelegenheit dar, da die benötigte Bandbreite der Server-Systeme respektive die Infrastruktur aufwändig in der Wartung ist und viel Geld kostet. Auch die rechtlichen Rahmenbedingungen – gerade bei illegal angebotenen Daten – spielen bei dieser Architektur eine wichtige Rolle, da Serverbetreiber meistens für den eigenen Inhalt zur Verantwortung gezogen werden können. Eine andere Architektur stellen die Peer-to-Peer-Netzwerke (P2P) dar. Diese basieren darauf, dass sehr viele Clients Dateien besitzen und diese untereinander mit mehreren Verbindungen über das Internet direkt austauschen. Wurden Daten aus dem P2P-Netzwerk geladen, bieten die Empfänger-Clients im Gegenzug oft ihre eigenen Dateien für andere an. Ein Client lädt also gleichzeitig von mehreren Clients Daten herunter, und im Gegenzug tauscht er diese auch mit mehreren Clients parallel aus. P2P-Netzwerke können hierbei von einem Server verwaltet werden (er verteilt dann nur die Informationen über die Clients untereinander, ohne selbst Dateien anbieten zu müssen), allerdings funktioniert das Ganze auch dezentral (ohne Server). Die Last und die Bandbreitenausnutzung werden somit von den Clients getragen und die Download-Server werden extrem entlastet. Bekannte Beispiele sind BitTorrent, eDonkey 2000 und das mittlerweile eingestellte Napster-Netzwerk, das diese Kommunikationsform einst populär machte. Gerade im Open-Source-Bereich stellen Betriebssystemhersteller über P2P-Netzwerke gerne ihre DVD-Images bereit und profitieren so von der aggregierten Bandbreite der Clients bei minimierter eigener Server-Belastung. Die mächtige P2P-Technologie hat jedoch leider – wenngleich zu Unrecht – einen schlechten Ruf, da auch viel illegaler Inhalt über diese P2P-Netzwerke verteilt wird. Zurzeit werden etwa 25 Prozent des gesamten Verkehrsaufkommens in Deutschland durch P2P verursacht.



Verteilung von FTP und NNTP(S) im Verhältnis zum Gesamtverkehr, Quelle IAS

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

Das Hauptproblem, weswegen die Benutzer mit dem geltenden Recht in Konflikt geraten, liegt im Anbieten von Inhalten, da bei P2P das Herunterladen auch wieder das gleichzeitige Anbieten eigener Dateien impliziert. Nachdem die Geschädigten von illegalen Downloads angefangen haben, Tausende Benutzer abmahnen zu lassen, und rechtliche Konsequenzen drohten, bildete sich eine neue Art der Datenverteilung, welche auf Clientseite zumindest einem einfachem Downloadprinzip folgte: Download und Upload mit Webbrowser-basierten Diensten, welche die Inhalte meist nur vom Anbieter zum Benutzer fließen lassen (Webfilesharing). Beispiele dafür sind Rapidshare und die kürzlich spektakulär geschlossene Plattform Megaupload. Viele Benutzer nutzen heutzutage die Webbrowser-basierten Dienste, da sie davon ausgehen, dass an dieser Stelle das Risiko, rechtlich belangt zu werden, niedriger liegt als bei P2P. Diese Services lassen sich als Cloud-Dienste klassifizieren, da kein Benutzer so genau weiß, wo die Daten liegen und auf welche Art und Weise sie gespeichert sind – geschweige denn wie sie verwaltet werden und wer darauf alles Zugriff hat. Auch ist die reale Verweildauer der hochgeladenen Daten nicht sicher und eine dauerhafte Speicherung nicht garantiert. Denn schließt ein Anbieter von heute auf morgen sein Geschäft (oder wird es geschlossen, wie im Falle von Megaupload), betrifft das natürlich auch alle dort abgelegten Daten.

Die ehemals viel genutzten Newsgroups oder das Usenet treten heute eher als Nebenerscheinung auf, wohingegen P2P-Verkehr oder die Webfilesharing-Plattformen (1-Klick-Hoster) immer noch stark vertreten sind und sogar immer mehr Nutzer anziehen. Die Anbieter rüsten im Bereich der Webfilesharing-Plattformen stark auf und investieren hohe Summen in Speicherplatz und ihre Internetanbindung. Jeder Abonnent mit einem Premiumzugang bringt Geld in die Kassen und erwirbt sich damit das Recht, oft mit unbegrenztem Traffic so viele Dateien herunterzuladen, wie er möchte. So hat beispielsweise Rapidshare nach eigenen Angaben eine Internetanbindung von 800 Gigabit/s für die rund 1.000 Server und mehrere Petabyte an Speicherplatz für die täglich hochgeladenen 400.000 Dateien. Bezüglich Wartung und Betriebsaufwand einer solch gigantischen

Infrastruktur geben die Betreiber keine Informationen. Auch über Zahlen zum Thema Marktanteile, aufkommendes Datenvolumen sowie registrierte Benutzer schweigen sich die 1-Klick-Hoster vollständig aus – sie möchten sich offenbar nicht in die Karten schauen lassen.

Downloads und Datenraten von 1-Klick-Hostern zu messen, gestaltet sich überdies für Dritte als herausfordernd, da es sich hierbei üblicherweise um normalen HTTP-Verkehr handelt und eine Analyse auf IP- oder sogar Inhaltsebene (Deep Packet Inspection) datenschutzrechtlich praktisch verboten ist. Statistische Analysen vom Institut für Internet-Sicherheit, die mit dem Internet-Analyse-System (IAS) aus eigenen Sonden in ganz Deutschland generiert wurden, geben zumindest einen Überblick über die Verteilung von FTP und NNTP(S) im Verhältnis zum Gesamtverkehr. So belegen NNTP und NNTPS etwa ein Prozent des Gesamtaufkommens, FTP kommt hingegen nur auf rund 0,06 Prozent. Der komplette HTTP(S)-Verkehr nimmt über 50 Prozent am Gesamtaufkommen in Deutschland ein. Hier sind auch die Downloads von 1-Klick-Hostern enthalten.

Ausblick

Wie viele andere Bereiche ist auch das Internet einer ständigen evolutionären Entwicklung ausgesetzt, in der sich die Kommunikationswege weiterentwickeln, neue hinzukommen oder alte an Bedeutung verlieren. Manche Konzepte sind innovative Neuentwicklungen, andere setzen auf bereits bestehenden Diensten beziehungsweise Technologien auf und nutzen diese auf völlig neue Art und Weise. Das jüngste Beispiel hierfür ist eine Technologie namens Share, die mit Hilfe der Nutzung des BitTorrent-Protokolls bestehenden Diensten wie Dropbox (sehr weit verbreiteter Speicherdienst in der Cloud) Konkurrenz machen soll. Bislang ist BitTorrents Share ebenfalls kostenlos. Im Gegensatz zu den bei vielen anderen Diensten geltenden Transfer- oder Speicherplatzbeschränkungen kennt Share hier derzeit keine Limits.

Beispiele für diese neue Art der Dateiweitergabe sind auch Sendoid, welches auf dem RTMFP-Protokoll (Real Time Media Flow Protocol) basiert, und Frenzy (Mac OS X). Sie agieren als eine Art neue soziale Plattform mit Dropbox-Technologie im Hintergrund und gewinnen durch ihre Einfachheit immer mehr Nutzer für sich. Das Thema Sicherheit scheint viele Nutzer bei solchen Diensten eher sekundär zu interessieren – Fragen dazu kommen in der Regel erst auf, wenn gravierende Mängel dafür gesorgt haben, dass Daten verloren oder entwendet worden sind – also wenn es bereits zu spät ist.

Auch die Entwickler solcher Dienste nehmen es mit der Sicherheit offenbar nicht ganz so genau. Diese Vermutung legt zumindest der Fall Dropbox nahe: Alle Shared-Pictures-Ordner sind standardmäßig für die Öffentlichkeit verfügbar und mit Google schnell aufgespürt. Die komplette Übernahme einer persönlichen Dropbox funktioniert ebenfalls, ist aber aus Sicht des Betreibers weder ein Mangel noch ein Sicherheitsproblem.

Das bereits erwähnte Sendoid ist in der Lage, riesige Datenmengen zwischen zwei Endpunkten im Internet oder im lokalen Netzwerk verschlüsselt zu übertragen. Der Benutzer muss sich hierbei keinerlei Gedanken über die Ziel-IP-Adresse, seine Firewall oder sonstige Dinge machen, da die Desktop-Applikation oder das Browser-Plugin von Sendoid alles automatisch für ihn regelt. Mag das dem privaten Nutzern ein Segen sein, ist es dem Systemadministrator üblicherweise ein Dorn im Auge: Daten lassen sich so verschlüsselt aus einem Firmennetzwerk heraus zu einem beliebigen Punkt übertragen, und das ganz ohne komplizierte Kraftakte und Tricks. Sendoid verbindet einige Funktionalitäten aus bereits bekannten Technologien wie dem BitTorrent-Protokoll, Dropbox und Webfilesharing und ermöglicht es, mit einem kleinen Programm (auf Adobe Air-Basis) Daten auch an unerfahrene Nutzer zu verteilen, die den Vorgang im Prinzip als normalen Browserdownload wahrnehmen.

Vorschau

In der Vergangenheit wurden Daten meist auf zwischenmenschlich anonyme Art und Weise unter den Benutzern ausgetauscht (etwa P2P). Die Frage nach dem Ursprung der Daten spielte eine untergeordnete Rolle, denn das Ausschlaggebende war, dass die angeforderten Daten früher oder später vollständig heruntergeladen wurden. Hierbei war keine Ressourcen-intensive Infrastruktur notwendig, um alle Teilnehmer oder Interessenten zu bedienen und zu verwalten. Der dadurch erzeugte Netzwerkverkehr zwischen den Clients war zwar sehr hoch (und ist es auch heute noch), allerdings ist hierzu keinesfalls unbedingt eine verwaltende Instanz (wie etwa zusätzliche kostspielige Server) notwendig, da die Clients (abhängig vom verwendeten Protokoll) in der Lage sind, selbstständig die richtigen Kommunikationspartner zu finden. Bei heutigen Cloud-Diensten wird hingegen eine größere Infrastruktur benötigt: In der Regel steckt hinter einem Dienst häufig sehr viel Speicherplatz mit einer sehr guten und auch dementsprechend teuren Internetanbindung. Über diese Dienste wurde der Datenaustausch persönlicher, da Daten meist im Bekanntenkreis verteilt werden und nicht mehr anonym, wie das bei P2P-Netzwerken der Fall ist. Anonymer Austausch ist hierbei jedoch im Sinne des unbekannten Gegenübers zu verstehen. Teil 2 des Artikels wird sich im kommenden Heft mit diesem Kommunikationsmodell „Datenaustausch mit bekannten Personen" beschäftigen. ■

Dominique Petersen ist wissenschaftlicher Mitarbeiter am Institut für Internet-Sicherheit der FH Gelsenkirchen und betreut dort seit Januar 2007 den Bereich der Internet-Frühwarnsysteme als Projektleiter.

Sebastian Barchnicki ist studentischer Mitarbeiter am Institut für Internet-Sicherheit der FH Gelsenkirchen und dort im Bereich Trusted Computing, Social Media und Mobile Security tätig.

Norbert Pohlmann, Informatikprofessor für Verteilte Systeme und Informationssicherheit sowie geschäftsführender Direktor des Instituts für Internet-Sicherheit an der FH Gelsenkirchen.