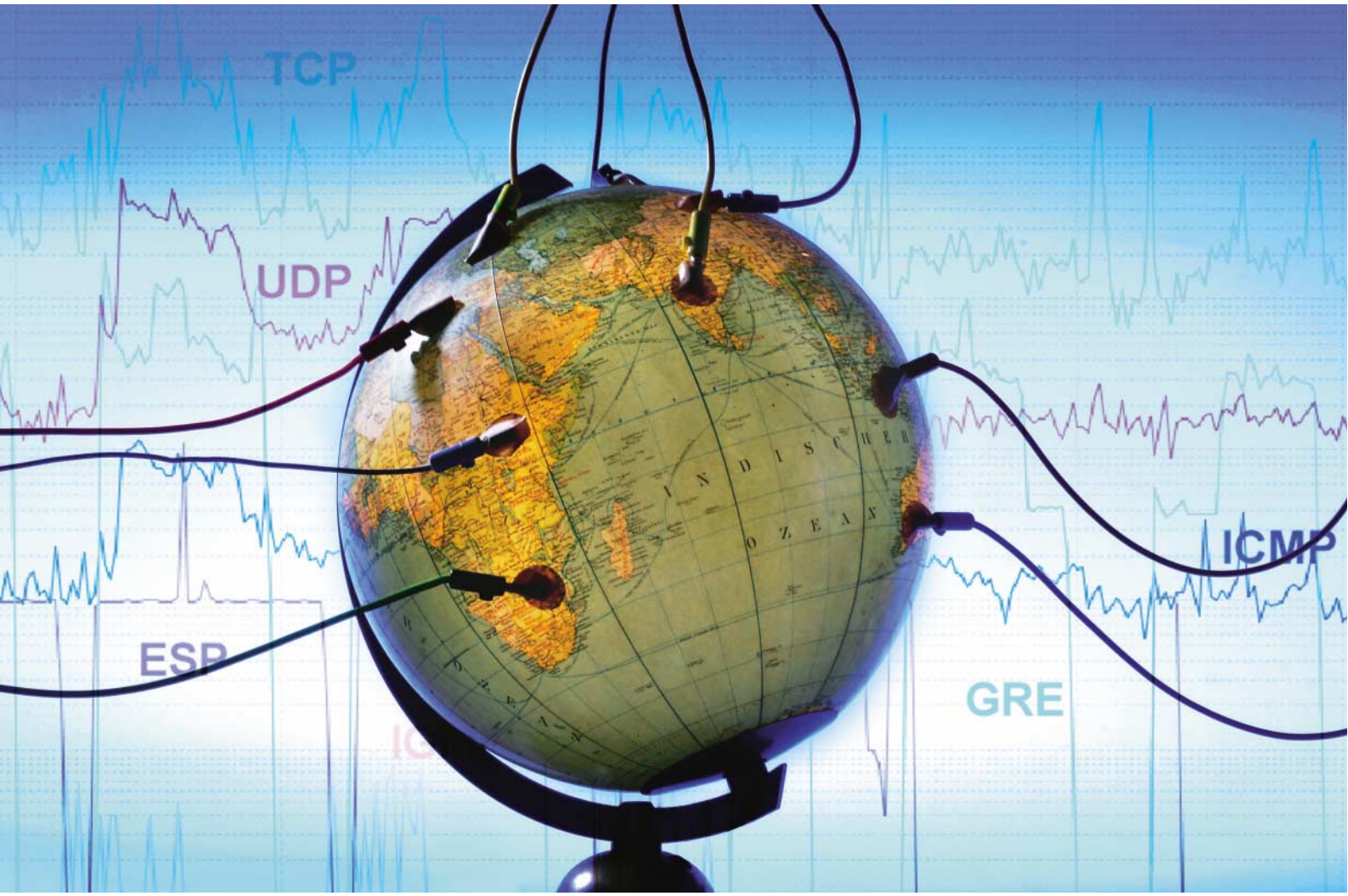


Die globale Sicht auf das Internet

Messverteilung

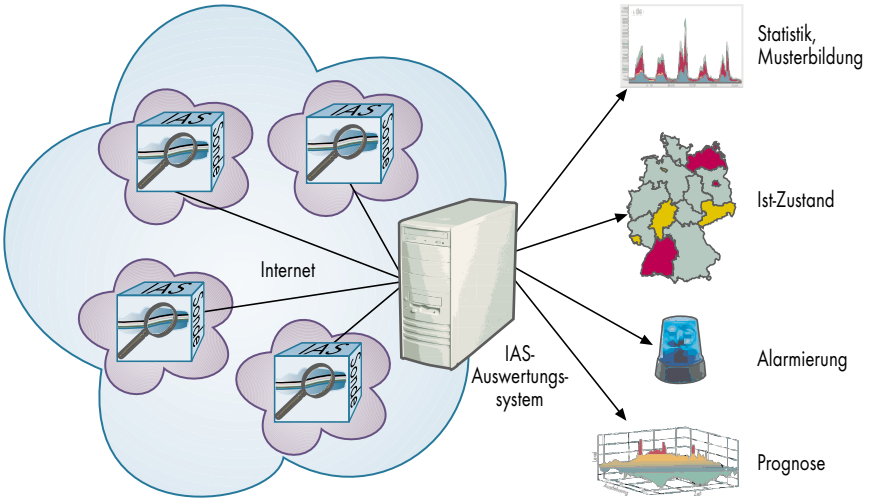
Norbert Pohlmann, Marcus Proest

Im Alltag von LAN-Administratoren funktionieren Kapazitätsplanung und Gefahrenabwehr schon leidlich. Doch im Internet, um einige Größenordnungen komplexer, fehlt es meist an korrelierten Messwerten über die laufende Daten-Verkehrslage. Ein verteiltes System von Sonden kann Abhilfe schaffen.

Autofahrer kennen die Situation: Man steht im Stau und sieht gerade einmal die Autokolonne vor und hinter sich. Es fehlt der Überblick, die Informationen über die Entstehung des Staus, seine Länge, die Position darin – und insbesondere darüber, wann er vorbei ist. Für Autofahrer gibt es Lösungen, die den Informationsmangel beseitigen. In Deutschland existiert ein engmaschiges Netz von Zählschleifen zur Erfassung der Verkehrssituation auf den Autobahnen. Stauinformationen stehen per Radio, SMS, Telefon und Internet bereit, und Navigationssysteme verarbeiten die Stauinformationen schon bei der Routenplanung. Globale Informationen befreien die Autofahrer von der lokalen, eingeengten Sicht, und helfen ihnen, frühzeitig Entscheidungen zu treffen, um alternative Wege zu finden.

Nicht anders verhält es sich mit der Sichtweise, die Netzbetreiber heute auf das Internet haben. In der Regel gibt es nur eine lokale Sicht, einen Überblick über die eigenen Netzsegmente und die Datenübertragung. Treten hier Engpässe auf, kann der Betreiber sie schnell und gezielt beheben. Zeichnet sich allerdings ab, dass sie nicht innerhalb der eigenen Handlungsdomäne auftreten oder dass die notwendige globale Sichtweise fehlt, wird es schwierig. Oft fehlt das Wissen um die Ursache, und deren Beseitigung erfordert die Hilfe Dritter.

© Copyright by Heise Zeitschriften Verlag GmbH & Co. KG. Veröffentlichung und Vervielfältigung nur mit Genehmigung des Heise Zeitschriften Verlags.



Aufgaben des Internet-Analyse-Systems (Abb. 1)

Eine globale Perspektive ist im Internet schwer zu erlangen, da sich niemand gerne in die Karten schauen lässt. Netzbetreiber behandeln interne Strukturen, Kommunikationsverbindungen und Topologien oft vertraulich. Kommunikationsdaten sind darüber hinaus im Prinzip datenschutzrelevant und die Datenübertragung so schnell, dass sie sich nicht immer „live" analysieren lässt. Eine Speicherung der Kommunikationsdaten für die Beobachtung langfristiger Entwicklungen erscheint unmöglich. Nicht zuletzt stellt sich die Frage, wer sich überhaupt für die Erstellung einer globalen Sichtweise verantwortlich fühlt.

iX-TRACT

- Die Vorgänge im Internet lassen sich mit Hilfe darin verteilter Sonden analysieren, die den Datenverkehr passiv mitschneiden.
- Eine Anonymisierung der Messdaten durch die Reduktion auf statistisch relevante Bestandteile wahrt den Datenschutz.
- Von großer Bedeutung für die rasche Erfassung der Daten durch Systemverantwortliche ist deren grafische Darstellung.
- Nicht nur die langfristige Planung von Infrastruktur, sondern auch die schnelle Abwehr von Gefahren ist mit Hilfe eines verteilten Analysesystems möglich.

Zurzeit realisiert das Institut für Internet-Sicherheit an der Fachhochschule Gelsenkirchen ein sondenbasiertes Internet-Analyse-System als Forschungs- und Entwicklungsprojekt in Zusammenarbeit mit dem Bundesamt für Sicherheit in der Informationstechnik (BSI). Es soll lokale und vor allem globale Sichtweisen erstellen und analysieren, um Frühwarnungen generieren zu können.

Schwerpunkte des Projekts sind eine datenschutzverträgliche Sammlung von Informationen sowie eine Optimierung des Ressourcenbedarfs bei deren Speicherung über große Zeiträume, die eine langfristige Analyse von Trends und Entwicklungen ermöglicht.

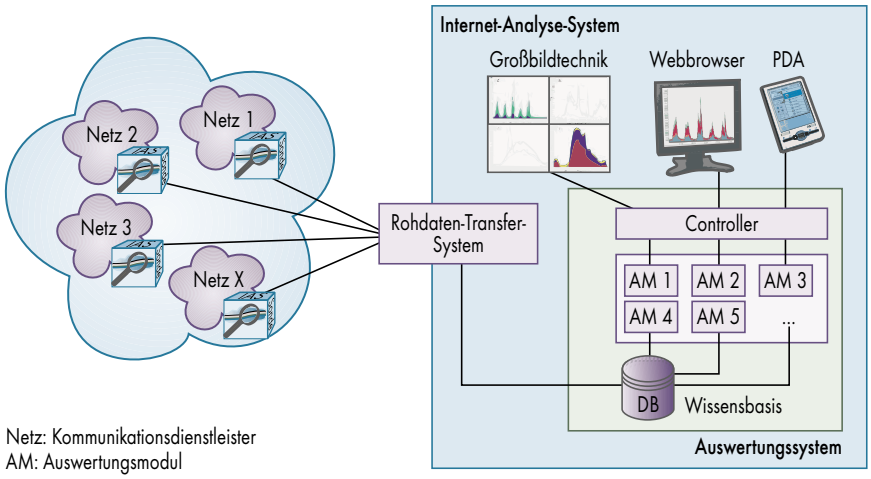
Das Internet-Analyse-System soll zum einen lokale Kommunikationsdaten in definierten Teilnetzen des Internet und zum anderen die Erstellung einer globalen Sichtweise auf das Internet durch die Zusammenführung der vielen lokalen Sichten abbilden. Es lässt sich in die vier Teilbereiche Musterbildung, Beschreibung des Ist-Zustandes, Alarmierung und Prognostizierung unterteilen.

Internet-Sonden zur Analyse

Als wesentliche Aufgabe kommt der Musterbildung eine umfassende Analyse und Interpretation der Kommunikationsparameter zu. Sie soll Trends und Zusammenhänge erkennen, Anomalien bei aktuellen Messwerten suchen und die Ursachen für die Zustandsänderungen analysieren und interpretieren. Es gilt herauszufinden, ob die Zustandsanomalien „natürlichen" Ursprungs sind, etwa durch die Einführung einer neuen Technik, oder ob ein mutwilliger Angriff zu Grunde liegt.

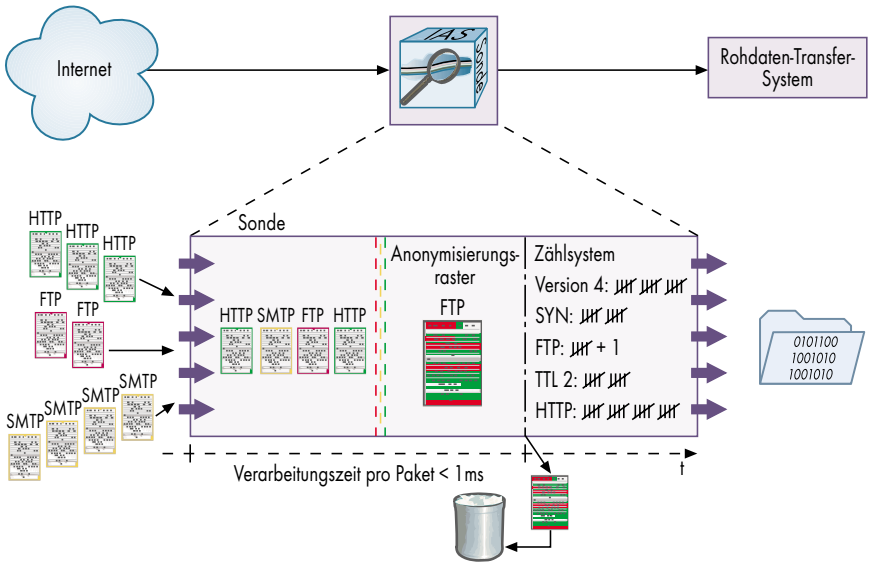
Ein Vergleich des aktuellen Zustands einer Kommunikationsleitung mit historischen Informationen (Wissensbasis) kann bei signifikanten Änderungen des Verkehrsaufkommens oder der Kommunikationsdaten zu einer Warnmeldung führen, gefolgt von Maßnahmen zum Schutz und Erhalt der Netzfunktion.

Analog zu einer Wetter- oder Staukarte lässt sich der Zustand des Internet darstellen, sodass die wichtigsten Parameter intuitiv erkennbar und für Prognosen über Zustandsänderungen geeignet sind. Angriffe und andere wichtige Veränderungen sollen frühzeitig auffallen, damit sich mögliche Schäden abwenden lassen [2].



Zusammenhänge zwischen den am Internet-Analyse-System beteiligten Komponenten (Abb. 2)

© Copyright by Heise Zeitschriften Verlag GmbH & Co. KG. Veröffentlichung und Vervielfältigung nur mit Genehmigung des Heise Zeitschriften Verlags.



Prinzip der Rohdatenerfassung: Eine Sonde schneidet Pakete von drei unterschiedlichen Anwendungssitzungen mit (Abb. 3).

Das Analyse-System besteht aus Sonden, die den Netzwerkverkehr an Kommunikationsleitungen unterschiedlicher Netze passiv abgreifen. Ein Auswertungssystem zeigt die Kommunikationsparameter übersichtlich an (Abb. 2).

Messpunkte im Netz verteilt

Für aussagekräftige Ergebnisse sind sehr viele Rohdaten erforderlich, praktisch ein Mitschnitt sämtlicher Kommunikationsparameter. Sie bestehen aus aggregierten Zählerständen, die das System in diversen Netzen sammelt (Abb. 2). In jedem der Teilnetze stellen Telekommunikations-Dienstleister (ISPs, Unternehmen, Hochschulen …) zum Beispiel einen Internetzugang für Endnutzer, Inhalte oder sonstige Dienste bereit. Sonden bei den Netzbetreibern liefern Rohdaten, die das eigens dafür spezifizierte Secure Raw Data Transfer Protocol (RDTPs) transportiert.

In jedem Netz sind eigene Analysen der von den Sonden gesammelten Daten möglich. Für eine globale, repräsentative Sichtweise des Internet sind Sonden in unterschiedlichen Netztypen erforderlich: bei global operierenden Providern ebenso wie bei Transit-, Content- und Endkundenprovidern sowie in größeren Unternehmensnetzen [1].

Die Sonden sollen aus dem Datenstrom Informationen extrahieren, die Aufschluss über den Zustand und die Nutzung der Kommunikationsstrecke und des oder der dahinter liegenden Netze geben – nicht nur für die Ermittlung von Trends, sondern insbesondere im Fall missbräuchlicher Nutzung oder bei einer Fehlkonfiguration. Alle Informationen muss der Sondenbetreiber auf ein notwendiges Minimum beschränken, damit sie sich auch rückwirkend über längere Zeiträume analysieren lassen – und nicht zuletzt, um dem Datenschutz Genüge zu tun.

Sonden im Datenstrom

Technisch betrachtet greifen die Sonden passiv, also ohne selbst den Datenverkehr zu beeinflussen, die Netzwerkverbindungen ab und zählen die Kommunikationsparameter der unterschiedlichen Protokolle. Die Zählergebnisse gelangen in festgelegten Zeitintervallen zum Rohdaten-Transfersystem und schließlich zur Darstellung, bei Bedarf auch zur Alarmierung (Abb. 3).

Von der Sonde kopierte Pakete gelangen durch mehrere Analyseklassen, die jeweils für ein bestimmtes Protokoll zuständig sind. Sie werten fest definierte Kommunikationsparameter in den Protokoll-Headern auf den unter-

Produkte für die verteilte IP-Analyse

Symantec (http://enterprisesecurity.symantec.de/) stellt das „DeepSight Threat Management System“, eine Lösung für Bedrohungsmanagement und Frühwarnungen zur Verfügung. Sie nutzt die von den teilnehmenden Benutzern zur Verfügung gestellten Logfiles aus Firewalls und Intrusion-Detection-Systemen sowie die Arbeitsergebnisse eines Analystenteams als zentrale Datenbasis für Reports, Analysen und Alarme. Die lassen sich im Web, per Fax, Mail, SMS et cetera abrufen.

Bei **DShield.org** (www.dshield.org) gibt es mit dem „Distributed Intrusion Detection System“ ein offenes, kostenloses System als Projekt des Internet Storm Centers (ISC) am SANS Institute. Als Datenquellen kommen viele bekannte Firewall- und IDS-Lösungen von registrierten und anonymen Benutzern zum Einsatz. An diesem Projekt kann jeder unbeschränkt teilnehmen. Den Upload der Log-Auszüge in die zentrale Datenbank besorgen ein Client-Programm per SMTP oder ein Web-Frontend.

Arbor Networks (www.arbornetworks.de) bietet mit seinen „Peakflow“ eine Netflow-basierte Management-Lösung, die Statistiken und Berichte erstellt und Netzwerkanomalien erkennt. Peakflow unterstützt die Routerkonfiguration, damit sich zum Beispiel DoS-Angriffe (Denial of Service) schnell unterdrücken lassen. Via „Fingerprint Sharing Alliance“ lassen sich einmal erkannte Angriffsmuster unter den Teilnehmern verbreiten.

Computer Associates' Produktreihe „Unicenter Netmaster“ (www.ca.com) misst die Netzwerkperformance und stellt sowohl Echtzeit- als auch historische Informationen zur Verfügung. Ein Alarm-Monitor informiert Administratoren über Auffälligkeiten in TCP/IP-Netzen.

Realtech (www.realtech.de) bietet mit „NetworkManager“ eine Lösung, die neben Geräten und Topologie den Netzwerkstatus über SNMP anzeigen kann. Als Zusatzmodul kann eine Erkennung von Anomalien und eine Anzeige von Statistiken über Netzwerkdaten verwendet werden, die ihre Daten aus dem NetFlow Protokoll bezieht. Die Lösung speichert Ereignisse und eignet sich zum Einsatz in Corporate LANs.

„Trace:Magic“ von **Synapse** (www.synapse.de) ermöglicht die Offline-Analyse von Netzwerkdaten, die beispielsweise *tcpdump* gesammelt hat. Die Software kann etliche Formate lesen und generiert aus den Daten Reports für die Erkennung von Konfigurationsfehlern oder Anomalien auch in großen, heterogenen Umgebungen.

© Copyright by Heise Zeitschriften Verlag GmbH & Co. KG. Veröffentlichung und Vervielfältigung nur mit Genehmigung des Heise Zeitschriften Verlags.

ID	Description		Count
131134	IP (Protocol Number 6)	:	18.854.151
131145	IP (Protocol Number 17)	:	1.123.149
327708	TCP (Flags: SYN)	:	334.435
327723	TCP (Flags: FIN/ACK)	:	480.697
327724	TCP (Flags: SYN/ACK)	:	275.779
545857	HTTP (Request Method POST)	:	2.026
545861	HTTP (Request Method GET)	:	293.616
545863	HTTP (Request Method HEAD)	:	18.992

Für die Analyse des Internet-Datenstroms genügt die Sammlung vollständig anonymer Protokoll-Zählerstände (Abb. 4).

schiedlichen Kommunikationsebenen aus, die nicht datenschutzrechtlich relevant sind. Die jeweils passenden Zähler halten ähnlich wie eine Strichliste die Häufigkeit bestimmter Header-Informationen fest. Bei den Rohdaten handelt es sich also um Aggregate von Zählern über einen definierten Zeitraum. Die Sonde löscht das Paket (in Abb. 3 ein FTP-Paket) sofort nach der Auswertung irreversibel und spurlos [3]. Die Wiederherstellung eines Paket-Kontextes oder auch nur eines Kommunikationsparameters ist weder möglich noch nötig.

In definierbaren Zeitabständen können die Sonden ihre Zählerstände (Rohdaten) an die Transfer-Systeme übertragen. Es handelt sich ausschließlich um vollständig anonyme Informationen (Abb. 4) in Form von Zählerständen und um eine unidirektionale Verbindung: Ein Verbindungsaufbau ist nur von Seiten der Sonde möglich, typischerweise alle fünf Minuten für etwa 20 Kilobyte.

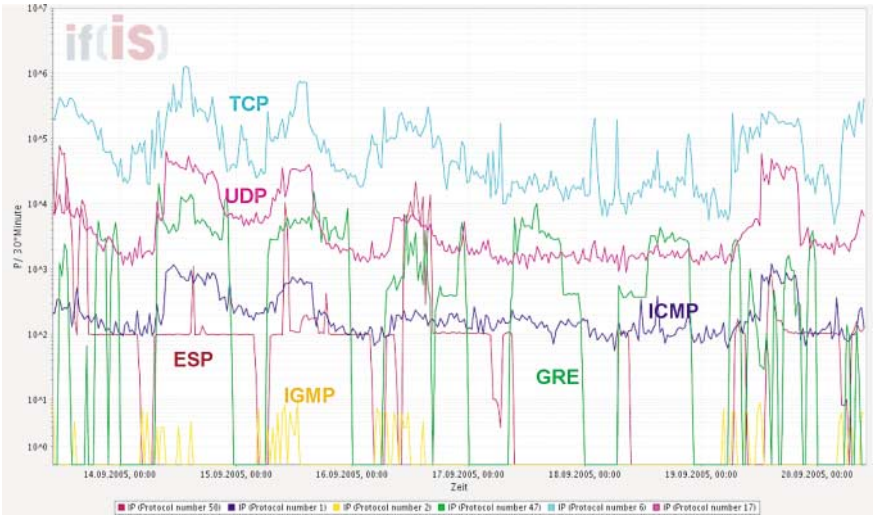
Da die Rohdaten nur eine statistische Formulierung der eigentlichen Kommunikationsdaten sind, würde es auch ausreichen, nicht jedes Paket zu betrachten, sondern zum Beispiel nur jedes zehnte – bei sehr viel Kommunikationsdaten eine pragmatische Lösung, die statistisch ähnlich signifikant ist.

Auswertung der Rohdaten

Die eigentliche Auswertung und Verarbeitung der gesammelten Informationen finden in diversen Analysemodulen (AM) des Auswertesystems statt, in Abbildung 2 mit „AM1“ bis „AM5“ gekennzeichnet. Die Module beziehen die Informationen ausschließlich aus der Wissensbasis (Rohdaten und Auswertungsergebnisse). Ziel der diversen Module ist das Erstellen von Profilen, Statistiken und Zusammenhängen, sowie das Erkennen von Schwellwert-Überschreitungen und die grafische Aufbereitung der Rohdaten und Auswertungsergebnisse.

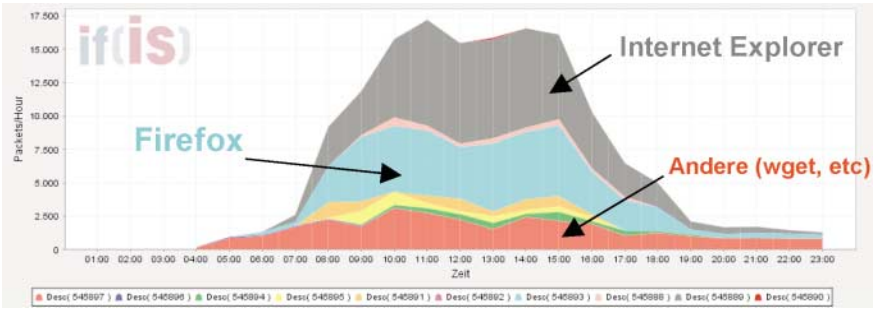
Dank der Anonymität der Rohdaten können Netzbetreiber sie miteinander austauschen oder an zentraler Stelle sammeln und als Basis für eine globale Sichtweise und Analysen eines IT-Frühwarnsystems verwenden.

Je nach Verwendung der Daten gibt es viele geeignete Methoden, die Ergebnisse der Internet-Analysen darzustellen (Abb. 2). Die Großbildtechnik dient der laufend aktualisierten Anzeige bestimmter Statistiken, Profile und aktueller Zustände. Ein intelligenter Client kann mit dem Auswertesystem



Protokoll-Aktivität in der Transportschicht über einen Zeitraum von einigen Tagen (Abb. 5)

© Copyright by Heise Zeitschriften Verlag GmbH & Co. KG. Veröffentlichung und Vervielfältigung nur mit Genehmigung des Heise Zeitschriften Verlags.



Datenverkehrsanteile von Web-Clients im Laufe eines Tages auf einer bestimmten Kommunikationsleitung (Abb. 6)

weiterreichende Analysen durchführen und Ergebnisse protokollieren. Außerdem lassen sich zum Beispiel mittels PDA Warnmeldungen des Systems mobil empfangen, damit sofort ein erster Überblick über die Gefahrenlage gegeben ist. Zurzeit dient dem Internet-Analyse-System ein Client als Frontend, der alle Zählerstände zu frei wählbaren Zeiträumen anzeigen kann. Weitere Funktionen sind per Plug-in-Technik ergänzbar.

Erfahrungen und Ergebnisse

Derzeit messen rund 300 000 Zähler die Parameter auf diversen Kommunikationsebenen. Die große Zahl verdeutlicht, wie komplex die Ergebnisse sein können und welche Bedeutung einer übersichtlichen Darstellung zukommt.

Abbildung 5 stellt die Verteilung der von der Transportschicht genutzten Protokolle über einen Zeitraum von mehreren Tagen auf einer bestimmten Kommunikationsleitung dar. Anomalien sind in dieser Form leicht zu erkennen. Aus der Vergangenheit sind Profil und Standardabweichung bekannt, sodass sich laufend Hinweise auf untypisches Verhalten gewinnen lassen – und auf die „normale“ Nutzung bestimmter Protokolle, etwa für die Kapazitätsplanung beim Einsatz von Virtual Private Networks. Auch Zusammenhänge zwischen Protokollen werden erkennbar: UDP scheint proportional zu TCP zu sein, was auf Abhängigkeiten zwischen HTTP und DNS zurückzuführen ist.

Auch die Anwendungen, die den Datenverkehr verursachen – etwa Webbrowser – lassen sich ermitteln. Abbildung 6 zeigt deutlich den Unterschied zwischen manueller (Internet Explorer, Firefox etc.) und automatischer Nutzung, etwa mit *wget*. Dabei handelt es sich keineswegs um die herkömmlichen Statistiken, die auf einen Webserver bezogen sind, sondern um solche, die sich auf eine Kommunikationsverbindung beziehen.

Fazit und Ausblick

Eine verteilte Erfassung des Internet-Datenverkehrs kann aufschlussreiche Informationen über den Netzzustand liefern – bis hin zu einer globalen Sichtweise auf Kommunikationstrends und Warnungen im Fall von Störungen und Angriffen. Je mehr Rohdaten es gibt, desto genauere Ergebnisse sind zu erwarten – insbesondere bei Prognosen.

Das Institut für Internet-Sicherheit an der FH Gelsenkirchen hat die zweite Entwicklungsphase abgeschlossen und das System auch am BSI in Betrieb genommen. Daneben läuft bereits die Erweiterung der bestehenden Protokolle, um zusätzliche Informationen zu erschließen, sowie der Auswertungsmodule und der Angriffserkennung. Hinzu kommen statistische Methoden und Datamining-Algorithmen, die den Evaluierungsprozess beschleunigen sollen.

Derzeit ist das Institut auf der Suche nach Partnern, die sich dem Verbund anschließen und weitere Sonden und Auswertungssysteme betreiben. Es gilt, das System gemeinsam flexibel und flächendeckend zu gestalten, um genau wie im Stau auf der Autobahn die globale Perspektive auf Strukturen und Gefahren zu schaffen. (un)

PROF. DR. NORBERT POHLMANN

ist Informatikprofessor für verteilte Systeme und Informationssicherheit sowie Leiter des Instituts für Internet-Sicherheit an der FH Gelsenkirchen (www.internet-sicherheit.de).

MARCUS PROEST

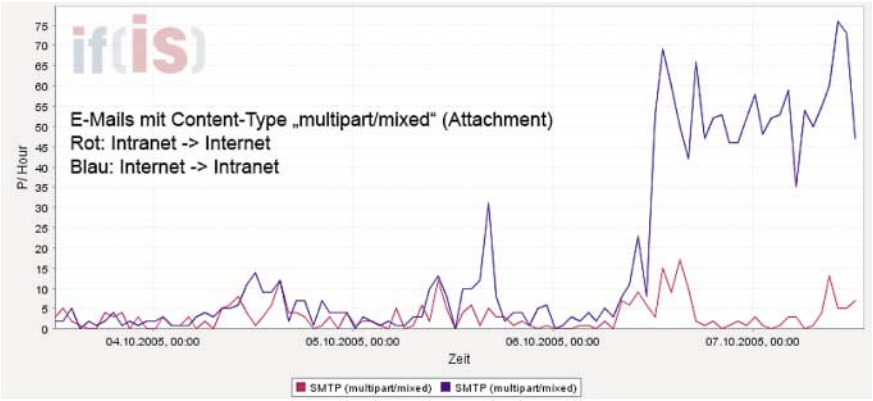
ist Diplom-Informatiker und für den Bereich Internet-Frühwarnsysteme im Institut für Internet-Sicherheit an der FH Gelsenkirchen verantwortlich.

Literatur

[1] Stefan Dierichs, Norbert Pohlmann; Netz-Deutschland: Zahl autonomer Systeme wächst stetig; *iX* 12/2005, S. 124

[2] Norbert Pohlmann; Internetstatistik; Proceedings of CIP Europe 2005; Hrsg.: B. M. Hämmerli, S. D. Wolthusen; Gesellschaft für Informatik, Bonn 2005

[3] Marcus Proest; Internet-Analyse – Ein Blick in die Dunkelheit; Konferenz: Internet-Sicherheit 2005; www.internet-sicherheit.de/center-berichte.html



Schlagartige Änderungen bestimmter Datenverkehrsanteile können auf ungewöhnliche Situationen im Internet hinweisen. Diese sprunghafte Zunahme an E-Mails mit dem Content-Type „multipart/mixed“ Anfang Oktober 2005 war auf die Sober.R-Welle zurückzuführen (Abb. 7).

© Copyright by Heise Zeitschriften Verlag GmbH & Co. KG. Veröffentlichung und Vervielfältigung nur mit Genehmigung des Heise Zeitschriften Verlags.