Internetstatistiken

Prof. Dr. Norbert Pohlmann

Institut für Internet-Sicherheit

Fachhochschule Gelsenkirchen
Fachbereich Informatik
Neidenburger Straße 43
45877 Gelsenkirchen

www.internet-sicherheit.de

norbert.pohlmann@informatik.fh-gelsenkirchen.de

Zusammenfassung

Die Aufgaben des Internet-Analyse-Systems lassen sich in die vier Teilbereiche Statistik, Beschreibung des Ist-Zustandes, Alarmsystem und Prognose einordnen. Hauptaufgabe der Statistik ist eine umfassende, statistische Analyse und Interpretation der Kommunikationsparameter des Internetverkehrs, mit dem Ziel, Profile, Technologietrends und Muster zu erkennen, und die unterschiedlichen Zustände des Internet darzustellen. Im nächsten Schritt werden Anomalien gesucht und die Ursachen für die Zustandsänderungen analysiert und interpretiert. Dabei ist es wichtig herauszufinden, ob die Zu-standsanomalien natürlichen Ursprungs sind (z.B. Technologietrend), oder ob ein mutwilliger Angriff zugrunde liegt. Falls eine mutwillige Attacke vorliegt, ist es von besonderer Bedeutung, die Muster zu finden, die den Angriff identifizieren.

Über exakte Kenntnis des aktuellen Zustands und Zuhilfenahme historischer Daten kann bei signifikanten Änderungen des Verkehrs eine Warnung ausgesprochen werden, aufgrund derer Maßnahmen zum Schutz und Erhalt der Funktionsfähigkeit des Internets ergriffen werden können.

Eine weitere wichtige Funktion ist die visuelle Darstellung des Internet-Zustands, analog zu einer Wetter-, oder Staukarte. Es soll eine intuitive Darstellung gefunden werden, mit der die wichtigsten Details auf den ersten Blick erkennbar sind.

Durch Untersuchung und Analyse der gefundenen Profile und Muster wird es möglich sein, Vorhersagen über Zustandsänderungen des Internets zu treffen. In Zukunft können geplante Angriffe erkannt und prognostiziert werden.

1 Einführung

Das Internet hat sich in den letzten Jahren zu einem allgegenwärtigen Medium entwickelt, das aus großen Bereichen der Wirtschaft, der Forschung und auch aus dem Privatleben nicht mehr wegzudenken ist. Deshalb ist die Analyse und Kenntnis des Mediums Internet von besonderer Bedeutung, um dessen Entwicklung bewerten und die zukünftige Funktionsweise aller enthaltenen Dienste gewährleisten zu können.

Die stetig steigende Bedeutung des Internets für unsere Gesellschaft macht es notwendig, die Zustände des Internets über die Grenzen der einzelnen Netzbetreiber hinweg zu analysieren bzw. zu kennen. Erst die genaue Kenntnis des Normalzustands macht es möglich, Abweichungen und Anomalien zu erkennen, die die Funktionalität des Internets beeinflussen. Mit Hilfe des Internet-Analyse-Systems, welches zurzeit als Forschungsprojekt des Instituts für Internet-Sicherheit realisiert wird, soll die Anwendung und Verteilung der im Internet genutzten Protokolle und Kommunikationsparameter herausgearbeitet und ausgewertet werden.

2 Aufgabe des Internet-Analyse-Systems

Die Aufgaben des Internet-Analyse-Systems lassen sich in die vier Teilbereiche Profilbildung, Beschreibung des Ist-Zustandes, Alarmsystem und Prognose einordnen. Hauptaufgabe des Profilbildungsbereichs ist eine umfassende Analyse und Interpretation der Kommunikationsparameter des Internetverkehrs, mit dem Ziel, Profile, Technologietrends, Zusammenhänge und Muster zu erkennen, die unterschiedliche Zustände und Sichtweisen des Internets darstellen. Auf dieser Wissensbasis werden Anomalien gesucht und die Ursachen für die Zustandsänderungen analysiert und interpretiert. Dabei ist es wichtig herauszufinden, ob die Zustandsanomalien natürlichen Ursprungs sind, wie z.B. Technologieveränderung, oder ob ein mutwilliger Angriff zugrunde liegt. Falls eine mutwillige Attacke vorliegt, werden die Muster identifiziert, die den Angriff charakterisieren.

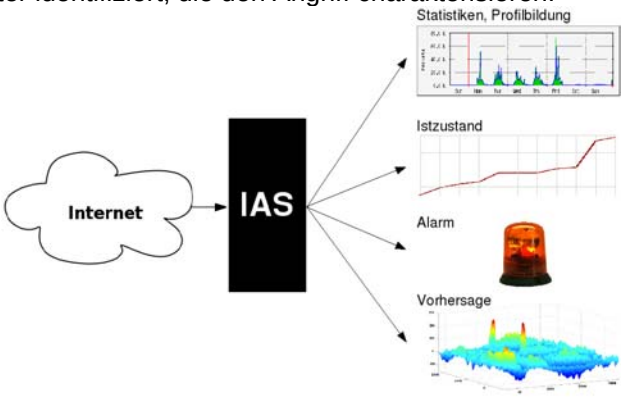


Abbildung 1: Ziele des Internet-Analyse-Systems

Über exakte Kenntnis des aktuellen Zustands (Kommunikationsparameter) und Zuhilfenahme historischer, das heißt zuvor erfasster Daten (Wissensbasis), kann bei signifikanten Änderungen des Verkehrsaufkommens oder der Kommunikationsdaten eine Warnmeldung generiert werden, aufgrund derer Maßnahmen zum Schutz und Erhalt der Funktionsfähigkeit des Internets ergriffen werden können.

Eine weitere wichtige Funktion ist die visuelle Darstellung des Internet-Zustands, analog zu einer Wetter-, oder Staukarte. Hier werden intuitive Darstellungen entwickelt, mit denen die wichtigsten Parameter auf den ersten Blick erkennbar sind.

Durch Untersuchung und Analyse der gefundenen Profile, Technologietrends, Zusammenhänge und Muster wird es durch einen Evolutionsprozess der gewonnenen Ergebnisse möglich sein, Vorhersagen über Zustandsänderungen des Internets zu treffen. Auf diese Weise können geplante Angriffe und wichtige Veränderungen erkannt und prognostiziert werden. Die folgende Abbildung zeigt eine Übersicht über die Ziele des Internet-Analyse-Systems.

3 Funktionsweise des Internet-Analyse-Systems

Das Internet-Analyse-System besteht aus Sonden, die passiv in verschiedenen Netzen Kommunikationsleitungen abgreifen und Kommunikationsparameter auf verschiedenen Ebenen zählt. Außerdem gibt es ein Auswertungssystem, welches die Kommunikationsdaten unter verschiedenen Gesichtspunkten auswertet und entsprechend anzeigt. Die Abb. 2 zeigt die Zusammenhänge zwischen den am System beteiligten Komponenten auf.

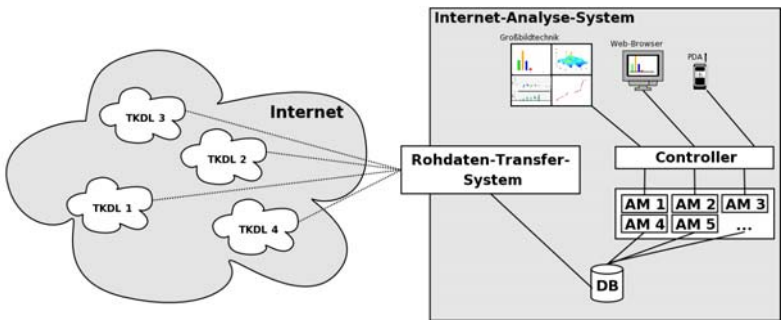


Abbildung 2: Komponenten des Internet Analyse Systems

TKDL: Telekommunikations-Dienstleister (ISPs, Unternehmen, Hochschulen, …);
AM: Auswertungsmodul

Damit das Internet-Analyse-System funktionieren kann, benötigt es eine möglichst große Menge an Rohdaten (Kommunikationsparameter). Auf diesen Rohdaten gründen sich alle Analysen, die das System durchführt. Sie bestehen aus aggregierten Zählerständen, die dem System von diversen Telekommunikations-Dienstleistern (TKDL) zur Verfügung gestellt werden (siehe Abb. 2). Auf der linken Seite der Grafik ist das Internet dargestellt, das aus einem Zusammenschluss zahlreicher Backbone-Netze besteht. In diesem Netz übernehmen Telekommunikations-Dienstleister (ISPs, Unternehmen, Hochschulen, …) die Aufgabe, einen Internetzugang für Endnutzer bereit zu stellen. Das Internet-Analyse-System (IAS) bezieht seine Rohdaten von den TKDL. Übermittelt werden die Daten über das eigens dafür spezifizierte Secure Raw Data Transfer Protocol (RDTPs). Die übertragenen Rohdaten bestehen aus Zählungen, die der TKDL über das Aufkommen bestimmter Kommunikationsparameter durchgeführt hat.

3.1 Funktionsweise der Sonden

Aufgabe der Sonden ist es, aus einem breiten Kommunikationsdatenstrom Informationen zu extrahieren, die Aufschluss über den Zustand und die Nutzung der Kommunikationsstrecke geben. Hierbei sollen alle Informationen erhalten bleiben, die nötig sind, um eine missbräuchliche Nutzung, eine Fehlkonfiguration, Trendentwicklungen oder eine Überlastung zu erkennen. Gleichermaßen soll aber die Menge der Informationen auf ein notwendiges Minimum beschränkt werden. Damit die Daten auch rückwirkend über längere Zeiträume betrachtet und analysiert werden können, wurde darauf geachtet, dass keine datenschutzrechtlich relevanten Informationen in den Extrakten der Sonden, den so genannten Rohdaten enthalten sind. Zusätzlich ermöglicht dies den Austausch der Daten zwischen den Telekommunikations-Dienstleistern, oder das Versenden an externe Stellen zu Zwecken weiterreichender Analysen.

3.2 *Prinzip der Rohdatenerfassung*

Abbildung 3 verdeutlicht das Prinzip der Rohdatenerfassung durch die Sonden. Sie gliedert sich in drei Teile. Links befindet sich schematisch das Internet. Es sind Pakete von drei unterschiedlichen Anwendungssitzungen dargestellt. Zusammengehörige HTTP Pakete, eine FTP Sitzung und eine SMTP Sitzung. In der Mitte der Grafik befindet sich das Sondensystem. Die Pakete der drei Anwendungen werden in zufälliger Reihenfolge nacheinander von der Sonde ausgewertet. Dann wird das Paket, das aktuell bearbeitet werden soll, durch das Anonymisierungsraster geschickt. Das Datenpaket, in der Grafik ein FTP Paket, wird danach sofort physikalisch, das heißt irreversibel und spurenlos gelöscht. Das Paket wird nun durch mehrere Analyseklassen geschleust, die jeweils für ein bestimmtes Protokoll zuständig sind und fest definierte Felder, die nicht datenschutzrechtlich relevanten sind, auswerten.

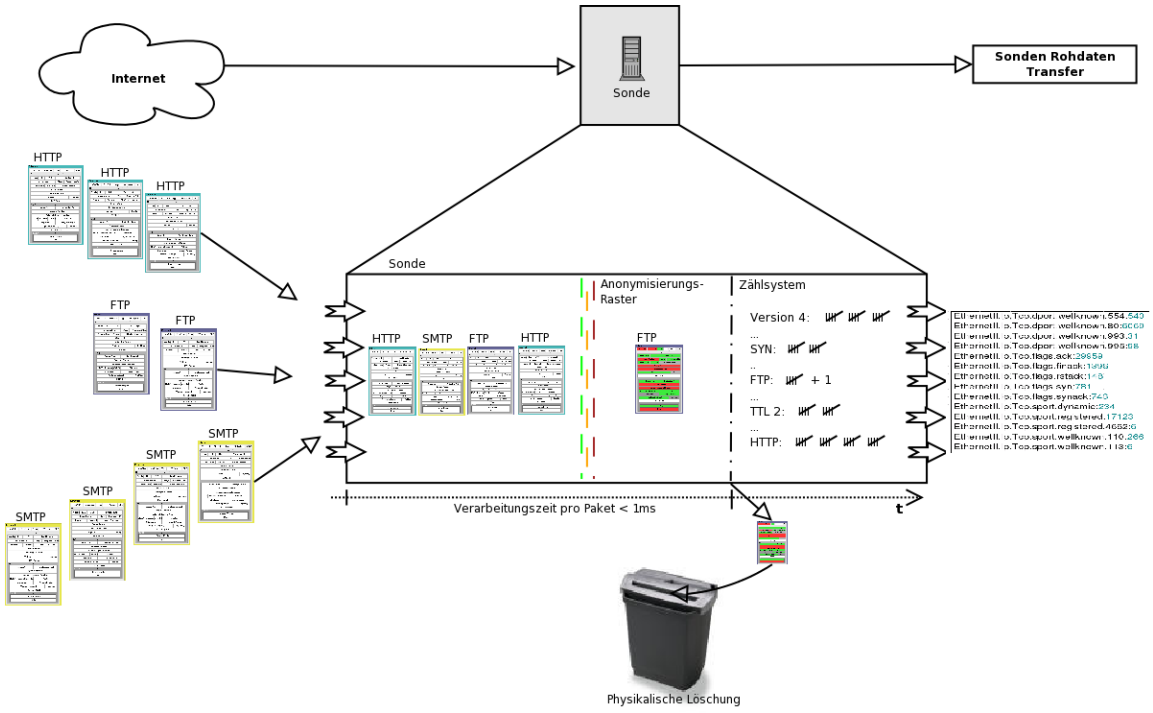


Abbildung 3: Funktionsweise der Sonde

Je nachdem, wie die Headerinformationen des Paketes aussehen, werden nun im Zählsystem zugeordnete Zähler inkrementiert. Ähnlich wie bei einer Strichliste wird die Häufigkeit bestimmter Steuerinformationen festgehalten. Beispielsweise wird in der Grafik die Häufigkeit von FTP Paketen ermittelt. Demzufolge wird der FTP Zähler um 1 inkrementiert. Bei den gespeicherten Analyseergebnissen handelt es sich also um Aggregate, das heißt um zusammenfassende Informationen vieler unterschiedlicher Pakete sehr vieler anonymer Kommunikationsentitäten.

```
EthernetII.Ip.Tcp.dport.wellknown.554:540
EthernetII.Ip.Tcp.dport.wellknown.80:6069
EthernetII.Ip.Tcp.dport.wellknown.993:31
EthernetII.Ip.Tcp.dport.wellknown.995:98
EthernetII.Ip.Tcp.flags.ack:29959
EthernetII.Ip.Tcp.flags.finack:1396
EthernetII.Ip.Tcp.flags.rstack:148
EthernetII.Ip.Tcp.flags.syn:781
EthernetII.Ip.Tcp.flags.synack:746
EthernetII.Ip.Tcp.sport.dynamic:234
EthernetII.Ip.Tcp.sport.registered:17123
EthernetII.Ip.Tcp.sport.registered.4662:6
EthernetII.Ip.Tcp.sport.wellknown.110:266
EthernetII.Ip.Tcp.sport.wellknown.113:6
```

Abbildung 4: Ausschnitt aus empfangenen Rohdaten

Anhand der gespeicherten Analyseergebnisse ist es aufgrund des Anonymisierungsrasters nicht mehr möglich, einzelne oder mehrere zusammenhängende Datenpakete zu rekonstruieren. Daraufhin können die Zählerstände (Rohdaten) von den Sonden zu analytischen Zwecken gespeichert, oder übertragen werden. Hierbei handelt es sich ausschließlich um die vollständig anonymen Sondenrohdaten, wie sie in Abbildung 4 zu sehen sind. Rechts hinter dem Doppelpunkt sind die Zähler für die links spezifizierten Headerinformationen. Jede Zeile steht für eine Zählung. Auf der linken Seite des Doppelpunktes steht die Zähl-Bedingung, rechts die Anzahl Pakete im Mess-Zeitraum. Zeile zwei der gezeigten Rohdaten bedeutet zum Beispiel, dass 6069 Pakete auf TCP-Zielport 80 (HTTP) eingegangen sind. Die Zählbedingungen und deren Kodierungen sind in einer versionierten XML-Datei spezifiziert. Wichtig ist es an dieser Stelle festzustellen, dass die übertragenen Daten und damit alle Daten, die das Internet-Analyse-System benötigt, aggregierte und vollständig anonyme Daten sind. Ein Personenbezug ist unter keinen Umständen herstellbar, ebenso ist es nicht möglich, Datenpakete einer Kommunikationssitzung und damit den Inhalt der Kommunikation zu rekonstruieren.

Das Rohdaten-Transfer-System fungiert als Server, mit dem sich die Zählsysteme der TKDL verbinden können, um ihre Rohdaten zu übertragen. Es handelt sich hier um eine unidirektionale Verbindungsmöglichkeit, das bedeutet, dass das Rohdaten-Transfer-System keine Möglichkeit hat, sich mit den Zählsystemen zu verbinden, ein Verbindungsaufbau ist nur von der anderen Seite aus möglich.

3.3 Auswertung der gesammelten Rohdaten

Die eigentliche Auswertung und Verarbeitung der gesammelten Informationen finden in diversen Analysemodulen statt. In Abbildung 2 sind diese durch „AM1“ – „AM5“ gekennzeichnet. Die Module beziehen die Rohdaten ausschließlich aus der Datenbank. Ein direkter Zugriff auf das Rohdaten-Transfer-System ist an dieser Stelle nicht möglich. Ziel der diversen Module ist das Erstellen von Profilen und Statistiken, sowie das Erkennen von Schwellwert-Überschreitungen und die grafische Aufbereitung der Rohdaten.

Da es sich bei den Rohdaten um vollständig anonyme Daten handelt, könnten sie zusätzlich zwischen verschiedenen Netzbetreibern ausgetauscht, oder an zentraler Stelle gesammelt werden, um sie als Basis für globale Analysen verwenden zu können.

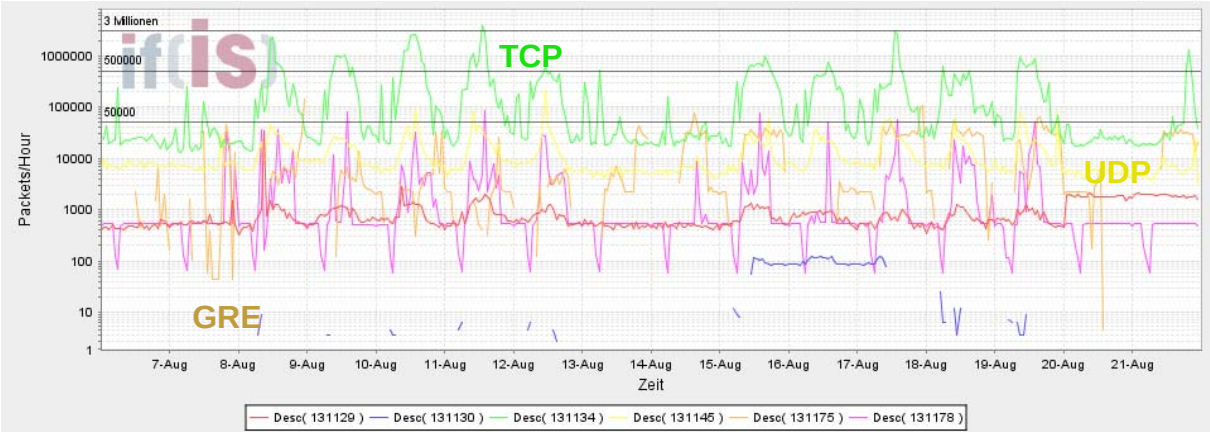
3.4 Benutzungsoberfläche des Internet-Analyse-Systems

Es sind viele Arten denkbar, das Internet-Analyse-System zu nutzen. In Abbildung 2 sind beispielhaft angeführt: eine Großbildtechnik, ein Web-Client, sowie ein PDA. Die Großbildtechnik dient der laufend aktualisierten Anzeige gewisser Statistiken und Profile. Über den Web-Client können weiterreichende Analysen durchgeführt werden und über den PDA können beispielsweise Warnmeldungen des Systems mobil empfangen werden, um sich sofort einen ersten Überblick über die Gefahrenlage verschaffen zu können.

4 Ergebnisse des Internet-Analyse-Systems

In diesem Kapitel werden zur Veranschaulichung einige Ergebnisse dargestellt, um eine Idee der Möglichkeiten zu vermitteln. Im Internet-Analyse-System werden zurzeit ca. 300.000 unterschiedliche Zähler von Kommunikationsparametern für die verschiedenen Kommunikationsebenen berücksichtigt. Diese große Zahl macht deutlich wie komplex die Ergebnisse sein können.

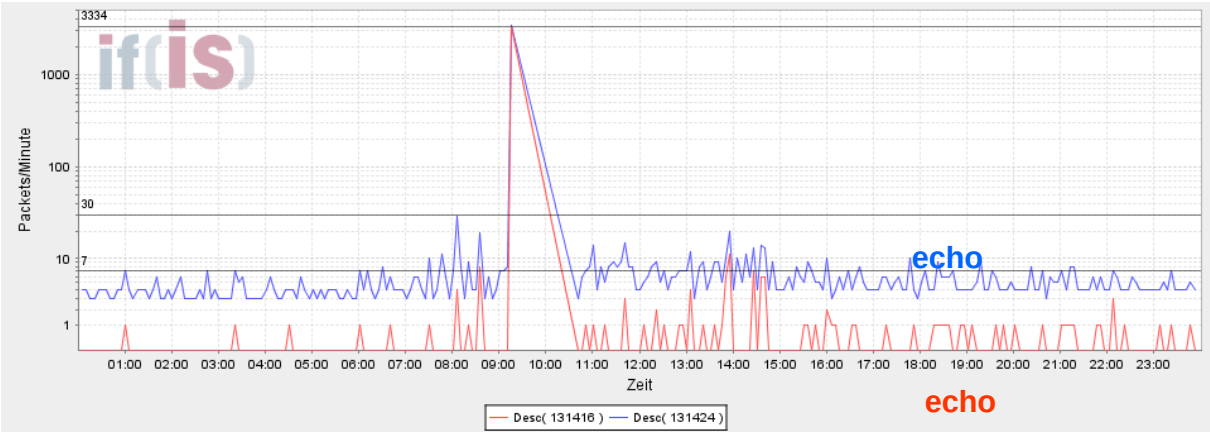
Transportprotokoll-Verteilung



TCP: Transmission Control Protocol
UDP: User Datagram Protocol
GRE: Generic Routing Encapsulation Protocol – VPN-Verbindung für das PPTP-Protokoll

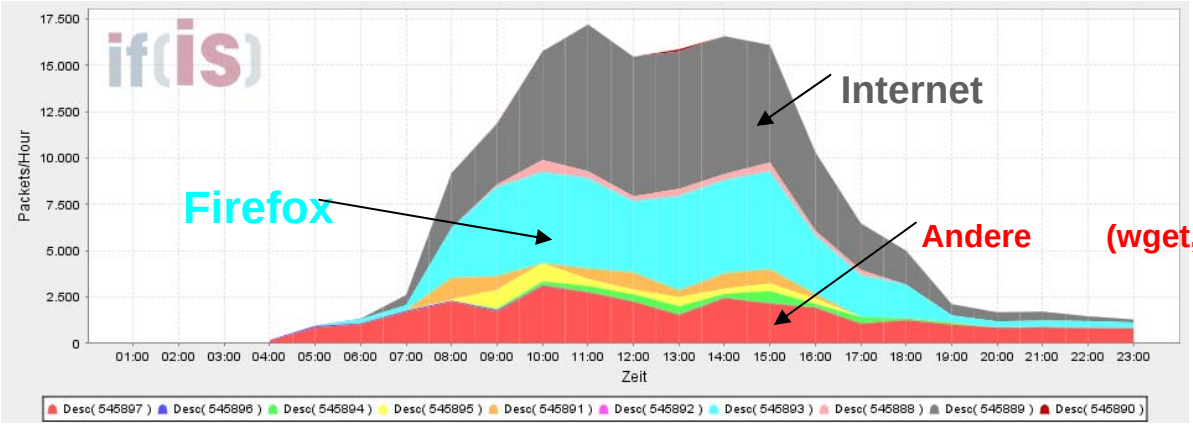
In dieser Graphik wir die Verteilung der genutzten Transportprotokolle über einen Zeitraum von mehreren Tagen einer bestimmten Kommunikationsleitung dargestellt. Aus dieser Graphik kann sehr leicht eine Anomalieerkennung erfolgen. Das Internet-Analyse-System kennt aus der Vergangenheit das Profil, die Standardabweichung und kann daraus sehr einfach einen Hinweis über untypisches Verhalen anzeigen.

„Ping of Death“ Angriff



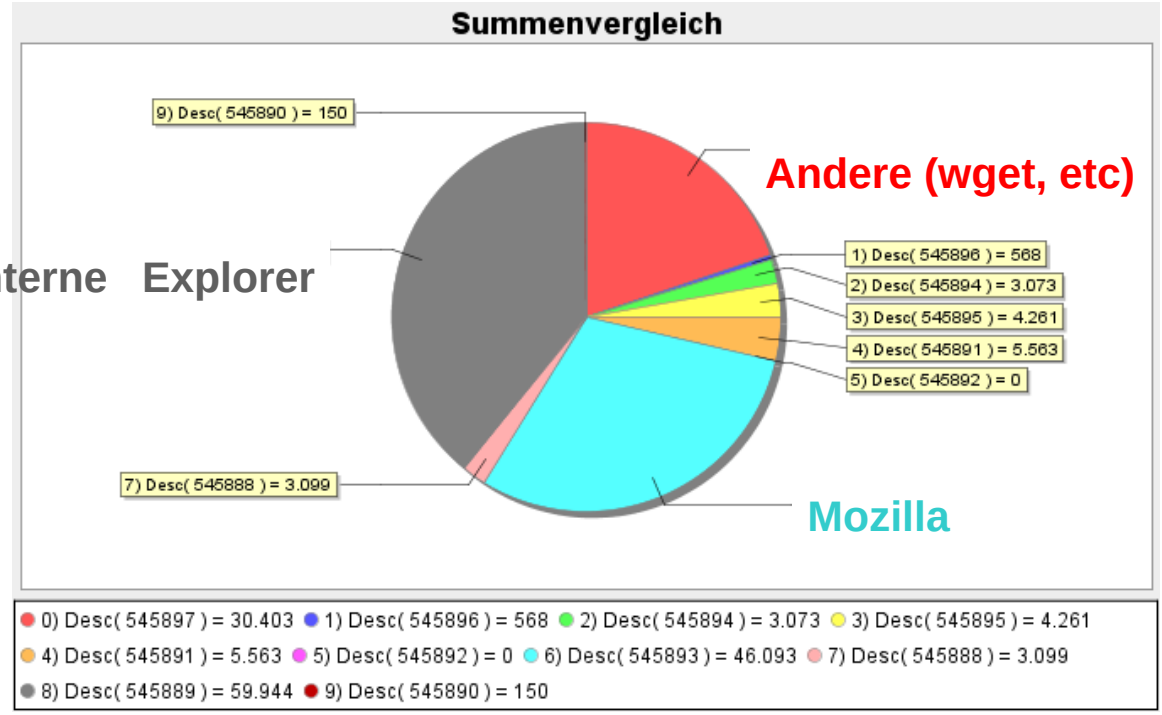
In dieser Graphik sind die Verteilung der ICMP-Pakete „echo request“ und „echo reply“ über einen Zeitraum von mehreren Tagen einer bestimmten Kommunikationsleitung dargestellt. In dieser Graphik können wir sehr leicht einen versuchten „Ping of Death“ Angriff erkennen. Jemand hat durch die schnelle Folge von „Ping“ Pakete (blaue Line) versucht einen Denial of Service Angriff (DoS-Angriff) durchzuführen. Da die Antworten des Angegriffenen fast synchron laufen (rote Line), hat der Angriff nicht seine volle Wirkung erzielt. Bei erfolgreichem Angriff würde das „eche reply“ (rote Line) deutlich zurück bleiben und auch sehr viel länger anhalten.

Browserverteilung (Technologie-Trend)



In dieser Graphik sind die Verteilung von unterschiedlichen Browsern über einen Zeitraum von einem Tag einer bestimmten Kommunikationsleitung dargestellt.

Hier können wir das Tagesprofil der unterschiedlichen Browser erkennen. Deutlich sieht man hier den Unterschied zwischen manueller Nutzung (z.B. Internet Explorer und Firefox) und automatischer Nutzung (z.B. wget) über den Tagesverlauf verteilt.



In dieser Graphik sehen wir den Summenvergleich, der über die Zeit wiederum den Technologietrend aufzeigt. Da diese Messung aus dem Hochschulbereich stammt ist Nutzung von „Mozilla Firefox“ höher als in anderen Netzen.

5 Fazit

Mit Hilfe des Internet-Analyse-Systems und des Sondensystems können kontinuierlich Rohdaten gewonnen werden, die den Internetverkehr statistisch widerspiegeln.

Durch die statistische Analyse der Rohdaten der Protokolle Vermittlungsebene und der Transportebene können weitreichende Informationen abgeleitet werden, die Aussagen über die Qualität von Verbindungen, die Nutzung von Protokollen und Diensten, sowie die Gefährdung durch Hacker und Würmer zulassen. Es lassen sich zudem Aussagen über die bisherige, aktuelle und künftige Nutzung von Diensten und Protokollen herleiten. Protokollanalysen auf Anwendungsebene geben Informationen über die Nutzung von wichtigen Diensten wie E-Mail und das World Wide Web, sowie die verwendeteten Versionen, Standards und Parameter wieder. Die Analyse von Rohdaten der kryptographischen Protokolle lässt Aussagen über die Verbreitung und Benutzung von Sicherheitstechnologien zu, sowie die verwendeten Algorithmen und PKIs.

Durch die Darstellung der Analyseergebnisse ist es möglich, einen Status des Internets darzustellen und Warnstufen im Fall von Problemen, wie beispielsweise infrastrukturellen Ausfällen oder Angriffen zu definieren. Bekannte Ansätze zur Prognose erlauben eine Analyse der Rohdaten und eine Vorhersage der Entwicklung von Trends in der Benutzung von Netzwerkdiensten.

Durch die kontinuierliche Analyse und das Einfliessen der gewonnenen Erfahrungen in den Analyseprozess steigt die Qualität der Ergebnisse stetig an.

Ein wichtiger Prozess ist die genaue Analyse der Rohdaten, um die Ergebnisse mit den Informationen anderer Netzanalysetools zu korrelieren. Mit den hergestellten Zusammenhängen können Muster erkannt werden, und somit Ergebnisse in Form von Statistiken, Mustern, Angriffsszenarien und Prognosen erstellt werden. Diese Ergebnisse können wiederum in den Sichtungs- und Auswertungsprozess einfliessen, um so immer früher immer genauere Ergebnisse zu erlangen. Hierbei ist der Umfang der Rohdatenbasis von entscheidender Bedeutung: Je mehr Rohdaten analysiert werden können, sowohl in Bezug auf die Anzahl der unterschiedlichen Sonden und deren Positionierung, als auch in Bezug auf die Zeitspanne der Daten, desto genauere Ergebnisse sind zu erwarten. Insbesondere gilt dies für die Vorhersageanalysen, da Algorithmen, die geeignet sind entsprechende Analysen vorzunehmen, eine besonders große Initialdatenmenge benötigen.

Das Internet ist seiner Natur nach äusserst flexibel, komplex und täglichen Veränderungen unterworfen. Um Statistiken über das Internet auch für beliebige zukünftige Technologien zur Verfügung stellen zu können, muss das Internet-Analyse-System in seiner grundlegenden Form ebenfalls absolut flexibel sein.

Damit das Internet-Analyse-System eine repräsentative Sicht auf das Internet gewährleisten kann, ist es sehr wichtig, dass die Standorte der Sonden adäquat gewählt werden. Die Internetkommunikation in Deutschland findet im Wesentlichen über Backbone Netze von Providern, Hochschulen und großen Firmen statt. Die sinnvolle Platzierung der Sonden an ausgesuchten Stellen dieser Netze ist von großer Bedeutung. Nur so kann ein repräsentatives statistisches Ergebnis erreicht werden.

6 Literatur

M. Proest: „Internet-Analyse - Ein Blick in die Dunkelheit“, Konferenz: Internet-Sicherheit 2005; http://www.internet-sicherheit.de/center-berichte.html

C. Dietrich, N. Pohlmann: „eMail-Verlässlichkeit – Verbreitung und Evaluation", in "DACH Security 2005", Hrsg.: Patrick Horster, syssec Verlag, 2005

N. Pohlmann: "Firewall-Systeme - Sicherheit für Internet und Intranet, E-Mail-Security, Virtual Private Network, Intrution Detection-System, Personal Firewalls", 5. aktualisierte und erweiterte Auflage, ISBN 3-8266-0988-3; MITP-Verlag, Bonn 2003