JOHANNES KEPLER UNIVERSITÄT LINZ | JKU

TNF

Technisch-Naturwissenschaftliche Fakultät

Cypherpunk Anonymous Remailer focusing on Natural Language Processing

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Masterstudium

NETZWERKE UND SICHERHEIT

Eingereicht von:
Robert Söllner, BSc.

Angefertigt am:
Institut für Netzwerke und Sicherheit

Beurteilung:
Assoz.Prof. Mag. Dipl.Ing. Dr. Michael Sonntag

Linz, Juli 2015

Zusammenfassung

Diese Masterarbeit beschäftigt sich mit der Entwicklung eines Cypherpunk Anonymous Remailers. Dieser dient der vollautomatischen Anonymisierung von Email Nachrichten. Der Anonymous Remailer soll erweiterten Schutz gegen die Analyse der Zuordnung von eingehenden zu ausgehenden (anonymisierten) E-Mails bieten. Beispielsweise wird die zeitliche Abfolge der Nachrichten verändert. Des Weiteren wird die Größe der Nachrichten durch zusätzlich erzeugte Header und zufällig angehängten Text beeinflusst. Es wird auf die Bestandteile dieser Software eingegangen und dabei werden die relevanten Techniken betrachtet.

Dabei wird zusätzlich ein besonderer Fokus auf Natural-Language-Processing gelegt. Relevante Objekte, z.B. Namen, Orte, Organisationen, usw. sollen im E-Mail Text erkannt und klassifiziert werden. Anschließend sollten diese, sofern gewünscht, durch generische Beschreibungen ersetzt werden. Diese generischen Beschreibungstexte dienen dazu den Lesefluss des Empfängers nicht zu stören. Damit keine Lücke im Text entsteht wird beispielsweise die Information, dass ein Personenname an einer Stelle im Text vorkam, durch den Tag [PERSON] symbolisiert. Dies ist für den Kontext des Textes wichtig.

Des Weiteren wird auf das Thema IT-Sicherheit und die Schutzziele, die bei der Konstruktion sicherer IT-Systeme definiert werden, eingegangen. Es wird anhand von Beispielen veranschaulicht, dass die Schutzziele Vertraulichkeit, Integrität, Verfügbarkeit, Zuordenbarkeit, (Rechts-)Verbindlichkeit, Anonymisierung und Pseudomisierung in einem Spannungsverhältnis zueinander stehen. Es kommt daher immer auf den gewünschten Einsatzweck eines Systems an, welche und in welchem Ausmaß die einzelnen Schutzziele relevant sind.

Abstract

This master thesis shows how to construct a Cypherpunk Anonymous Remailer. It is used to automatically anonymize email messages. The Anonymous Remailer should implement advanced techniques to impede the possibility of finding the correlation between received and sent mail by observing the in- and output. Therefore the email messages should be randomly delayed. Furthermore the size of the messages should be changed by adding generated headers and adding generated random text at the end of the e-mail body. The components of this software will be described and the relevant techniques are introduced.

An additional focus is set on the field of Natural-Language-Processing. Relevant objects, e.g. names, locations, organizations have to be found and classified in the e-mail text. The tags can then be replaced with generic descriptions if required. These anonymized tags are inserted to not disturb the natural flow of reading. For example to symbolize that a person name was deleted in the text a [PERSON] tag will be inserted at this position in the text. If you just leave them out the reader would probably miss the context of the text.

Furthermore the important topic of Information Security and the Attributes of Information Security that are essential for the construction of secure IT-systems are presented and described. Examples show the tension between the different Information Security Attributes, confidentiality, integrity, availability, accountability, non-repudiation and anonymization and pseudonymization. It always depends on the intended use of a system what and of which weight the Information Security Attributes are relevant.

Danksagung

Ich möchte allen Personen danken, die mich bei der Erstellung meiner Masterarbeit unterstützt haben. Insbesondere bei Herrn Assoz.Prof. Mag. Dipl.-Ing. Dr. Michael Sonntag für die konstante Betreuung von Institutsseite aus und für die Unterstützung bei den Masterarbeitsseminaren. Bei Herrn Oberrat Dipl.-Ing. Dr. Rudolf Hörmanseder für die technische Unterstützung.

Herzlich bedanken möchte ich mich ebenfalls bei meiner Familie für die kontinuierliche Unterstützung und das Verständnis, sowie bei meinen Freunden für das Korrekturlesen.

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Einleitung

1.1. Motivation

Ziel dieser Masterarbeit ist es einen Cypherpunk Anonymous Remailer (Type I) zu entwickeln. Dieser soll zusätzlich zu den technischen Übermittlungsdaten auch den Inhalt der E-Mail Nachrichten analysieren um den textuellen Inhalt der E-Mail Nachrichten zu anonymisieren. Dies ist ein neuer Ansatz, der einen Fokus auf das in den letzten Jahren stark fortgeschrittene Feld des Natural Language Processing legt. Natural Language Processing beschäftigt sich mit der algorithmischen Analyse von Text. Für den Remailer ist vor allem das Subfeld der „Name Recognition“ von Interesse. Hierbei wird versucht Objekte z.B. Personennamen, Orte, E-Mail Adressen, Organisationen, in einem Text zu erkennen und deren semantischen Typ zu bestimmen. Auf diese Art und Weise lässt sich mithilfe von gezielten Ersetzungen ein personalisierter Text in einen generischen (nicht mehr zuordenbaren) Text umwandeln. Dieser Vorgang wird „anonymisieren“ genannt.

Im Folgenden wird erklärt was ein Anonymous Remailer ist und welche Arten von Anonymous Remailern es gibt. Des Weiteren werden die netzwerktechnischen Grundlagen mit dem ISO/OSI-Schichtenmodell und TCP/IP-Schichtenmodell erklärt und auf Anonymisierungsmöglichkeiten geprüft. Die für den E-Mail Verkehr notwendigen Applikationsprotokolle und Standards werden beschrieben und auf ihre Sicherheit und Anonymisierbarkeit wird eingegangen. Es wird eine konkrete Software Implementierung eines Anonymous Remailers (Typ I) vorgestellt und anhand von Beispielen und Tests erklärt was in der Praxis die Schwierigkeiten und deren Lösungsmöglichkeiten der Anonymisierung sind. Insbesondere wird auf das Feld des „Natural Language Processings“ eingegangen. Anhand einer statistischen Auswertung werden unterschiedliche Classifier des Stanford Named Entity Recognizer (NER) Frameworks verglichen und der Eignung für die E-Mail Anonymisierung anhand praxisrelevanter E-Mail Texte überprüft.

Wie auch andere Anonymisierungsservices wie beispielsweise das “Tor network” werden Anonymous Remailer vor allem für den Schutz der Privatsphäre eingesetzt. Dieser ist vor allem für Personengruppen relevant, die mit Verfolgung rechnen müssen, falls Ihre Identität bekannt wird. Beispielsweise begeben sich Menschenrechtsaktivisten in große Gefahr wenn sie Missstände in Ländern aufdecken, die Kritiker unterdrücken und verfolgen. Mitunter wollen auch Mitarbeiter von Unternehmen illegale Missstände im jeweiligen Unternehmen aufdecken. Da es für die betroffenen Unternehmen in solchen Fälle um große Geldsummen gehen kann, liegt es im Interesse des jeweiligen Mitarbeiters anonym bleiben zu wollen. Zusammengefasst kann man sagen Behörden, Journalisten und generell die Öffentlichkeit können von Anonymous Remailern profitieren, da sie es Informanten ermöglichen anonym zu

bleiben aber dennoch wichtige Informationen weiter zu geben, die andernfalls vermutlich nie öffentlich gemacht werden würden. Je nach Konfiguration des Anonymous Remailers ist zu beachten, dass Personennamen, Orte und Organisationsnamen auch aus dem Text des E-Mails gefiltert werden um generelle Anonymität zu gewährleisten. Dies muss bei der Verwendung von Anonymous Remailern berücksichtigt werden.

1.2. Begriffsdefinition: Anonymous Remailer

Ein Anonymous Remailer ist eine Software, die E-Mail Nachrichten empfängt und alle Daten, die auf den ursprünglichen Absender schließen lassen, entfernt. Dieser Vorgang wird auch als „Entpersonalisierung“ bezeichnet. In der empfangenen Nachricht muss der eigentliche Empfänger der Nachricht angegeben sein, da ein Remailer nur eine Zwischenstation ist. Der Anonymous Remailer leitet die Nachricht anschließend an den/die eigentlichen Empfänger weiter. Für den Empfänger gibt es keine Möglichkeit auf die Nachricht zu antworten, da für ihn nur der Anonymous Remailer als Absender ersichtlich ist.
Es ist nicht genau bekannt wann der erste Anonymous Remailer eingesetzt wurde. Cypherpunk Remailer (Type I) wurden in den 90er Jahren entwickelt. Damit sollte das Recht der freien Meinungsäußerung durchgesetzt werden. Personen die unter politischer, kultureller, ethnischer, religiöser oder irgendeiner anderen Form an Verfolgung leiden bzw. davon bedroht werden können einen Anonymous Remailer verwenden, um Ihre Nachrichten anonym zu verbreiten. Im positiven Sinn kann ein Anonymous Remailer daher als Instrument zur Stärkung der Rede- und Meinungsfreiheit im Internet gesehen werden. Leider zieht Anonymität auch Kriminelle an, die versuchen Anonymous Remailer für unlautere Zwecke wie z.B. Erpressung oder SPAM zu missbrauchen. Zumindest die SPAM Problematik kann mit technischen Schutzmaßnahmen wie z.B. der Limitierung empfangener Nachrichten pro Zeiteinheit vom selben Absender oder dem Einsatz von SPAM-Filtern eingedämmt werden.

Um Nachrichten noch besser vor Rückverfolgbarkeit schützen zu können, lassen sich auch mehrere/unterschiedliche Remailer miteinander kombinieren. Diese Methode ist besonders effektiv, wenn Nachrichten jeweils mit dem Public Key des nächsten Remailers verschlüsselt werden und ineinander verschachtelt verschickt werden. So kennt jeder Remailer nur jeweils den Remailer von dem er die Nachricht bekommen hat und den Remailer an den die Nachricht weiter verschickt wird. Nur der erste Remailer in der Kette kennt den ursprünglichen Absender und der letzte Remailer in der Kette kennt den eigentlichen Empfänger.

Die folgende Abbildung veranschaulicht das „Wrap“ Prinzip der Verkettung mehrerer Anonymous Remailer:

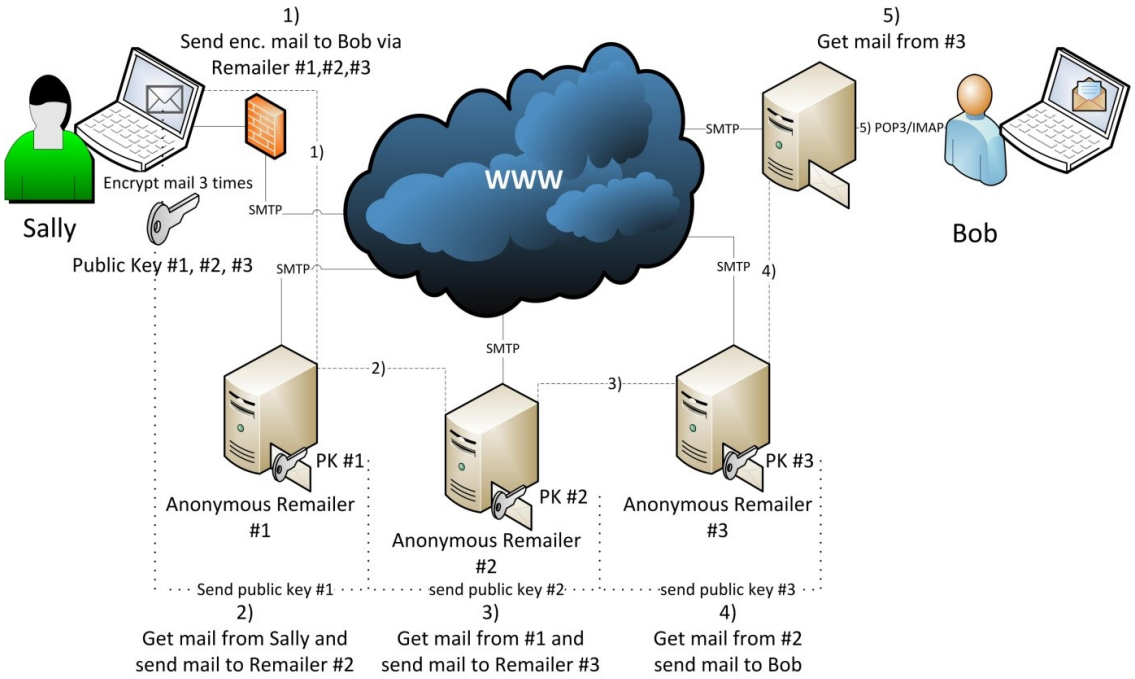

Abbildung 1: Visualisierung "Wrap" Prinzip

1.3. Arten von (Anonymous-) Remailern

Die Vorstufe der Anonymous Remailer sind die Pseudo-Anonymous Remailer. Ein Pseudo-Anonymous Remailer ersetzt E-Mail Adressen in E-Mail Nachrichten mit generierten Pseudonymen. Sofern dieses Pseudonym-zu-E-Mail-Adresse Mapping gespeichert bleibt, ist mit dieser Remailer Art eine wechselseitige Kommunikation zwischen Sender um Empfänger möglich, ohne das die Beteiligten die echte E-Mail Adresse ihres Kommunikationspartners kennen müssen. Diese Remailer Art unterbindet aber nicht, dass z.B. mittels Gerichtsbeschluss die Herausgabe der E-Mail Adresse unter Vorlage des Pseudonyms verlangt werden kann. Eine echte Anonymität ist daher bei den Pseudo-Anonymous Remailern nicht gegeben. Anonymous Remailer hingegen verzichten auf die wechselseitige Kommunikation. Ziel ist es die Nachricht so zu anonymisieren, dass sie nicht mehr dem eigentlichen Absender zugeordnet werden kann. Man unterscheidet drei verschiedene Grundtypen der Anonymous Remailer:

- Cypherpunk Anonymous Remailer (Type I)
Empfängt (verschlüsselte) E-Mail Nachrichten, entschlüsselt diese und entfernt möglichst jegliche Information, die auf den ursprünglichen Absender schließen lässt. Anschließend werden die Nachrichten verschlüsselt an den eigentlichen Empfänger übermittelt. Es besteht die Möglichkeit den Inhalt der E-Mail Nachrichten mehrfach zu verschlüsseln („Wrap“ oder „Zwiebel“-Prinzip) und sie durch eine Kette von Cypherpunk Remailern zu senden (Server müssen vom Sender selbst ausgewählt werden - kein Automatismus).

- Mixmaster Anonymous Remailer (Type II)

Das Prinzip ist hier ähnlich wie bei Remailer Type I, jedoch müssen die E-Mail Nachrichten vor dem Senden zum Remailer in einheitlich große Pakete aufgeteilt werden um eine Zuordnung zum Sender anhand der Nachrichten-Größe zu verunmöglichen. Daher wird ein eigenes Mailprogramm benötigt, welches diese gleichmäßige Aufteilung durchführt. Es werden diverse Techniken wie z.B. Verschlüsselung und Umsortierung der Reihenfolge der Nachrichten eingesetzt, um eine Zuordnung von eingehenden und ausgehenden Nachrichten zu verhindern.

- Mixminion Anonymous Remailer (Type III)

Das Prinzip stellt eine Erweiterung des Remailer Type II dar. Es wird ebenfalls ein eigenes Mailprogramm benötigt, welches die Nachrichten in einheitlich große Pakete aufteilt. Zusätzlich steht ein Netzwerk an Servern zur Verfügung. Die Kommunikation zwischen diesen Servern findet verschlüsselt statt. Die E-Mail Nachricht wird an einen Server des Remailer-Netzwerks geschickt - dieser wird „start mix“ genannt. Der Server schickt die Nachricht zufällig an einen anderen zugehörigen Server weiter. Die Nachricht wird nun von Server zu Server geschickt, bis der so genannte „exit mix“ erreicht ist. Jeder Server kennt jeweils nur den vorherigen Server und den nächsten Server. Die Nachrichten sollen dadurch nicht rückverfolgbar sein. [DDM03]

1.4. Stufen von Anonymität

Je nach Art und Einsatzgebiet des Systems ist es sinnvoll/erforderlich ein definiertes Level an Anonymität zu schaffen. Man kann dabei sechs Stufen von Anonymität auf dem Weg vom eindeutig identifizierbaren zum komplett anonymen User unterscheiden:

- **Super-Identifikation**

Bei der Super-Identifikation findet die Authentifizierung des Users außerhalb des eigentlichen Systems statt, sozusagen durch einen vertrauenswürdigen Dritten. Beispielsweise nach dem X.509 Standard für eine Public-Key-Infrastructure zum Erstellen vertrauenswürdiger Zertifikate. Das System muss dann nur noch die Gültigkeit des Zertifikats überprüfen.

- **Identifikation**

Das System identifiziert den User selbst durch Abfrage einen Geheimnisses. Das Geheimnis muss zuvor initial mit dem User vereinbart worden sein. Wenn das vom User mitgeteilte Geheimnis mit dem intern abgespeicherten Geheimnis übereinstimmt, so gilt der User als authentifiziert. Das Geheimnis ist normalerweise ein Passwort. Diese Methode wird bei einer Vielzahl an Applikationen verwendet.

- **Verborgene Identifikation (kontrollierte Pseudonyme)**

Das System authentifiziert den User und vergibt ein fix zugeordnetes Pseudonym für den User. In der Folge kann der User mit Hilfe des Pseudonyms personalisierte Dienste in Anspruch nehmen ohne seine wahre Identität Teilkomponenten des Systems zugänglich zu machen. Unter bestimmten Umständen kann das System das Pseudonym wieder in die reale Identität umwandeln. Ein Beispiel hierfür sind „Chiffre“ Anzeigen in Zeitungen.

- **Identifikation mittels Pseudonym (unkontrollierte Pseudonyme)**

Wenn der User das System zum ersten Mal verwendet wählt er ein Pseudonym und ein geheimes Passwort aus. Fortan verwendet er das Pseudonym und das Passwort für die Authentifizierung. Das System kennt die wahre Identität nicht, es kann den Benutzer nur dem jeweiligen Pseudonym zuordnen. Diese Methode wird beispielsweise bei einigen Internetforen verwendet. Sie kommt ebenfalls bei der Vorstufe der Anonymous Remailer, den Pseudo-Anonymous Remailern zum Einsatz.

- **Anonyme Identifikation**

Der User erhält Zugriff auf das System nachdem er sich durch ein Geheimnis (z.B. Passwort) als berechtigt ausgewiesen hat. Das System kennt weder die Identität des Users noch kann es User voneinander unterscheiden, die das gleiche Geheimnis kennen. Ein Beispiel hierfür sind Sparbücher, wie es sie früher gab. (Inzwischen ist eine Authentifizierung vorgeschrieben). Damals war es ausreichend das Passwort für das Sparbuch zu kennen, um darüber verfügen zu können.

- **Anonymität**

Das System verlangt keinerlei Authentifizierung oder Identifikation vom User. Es kann daher nicht zwischen Usern unterscheiden. Jeder User stellt eine eigenständige Entität dar, auch wenn es der gleiche User ist. Echte Anonymität ist im Internet schwer herzustellen, da die digitale Kommunikation Spuren hinterlässt die dazu verwendet werden können, zumindest zwischen unterschiedlichen und gleichen Usern zu unterscheiden. Beispielsweise die IP Adresse, MAC Adresse oder eine Applikationskennung sind meistens bei verschiedenen Benutzern unterschiedlich. Echte Anonymität lässt sich im Internet nur durch großen Aufwand herstellen. Die verschiedenen Arten von Anonymous Remailern drücken dieses Ziel aus, möglichst das Level „Anonymität“ für den User zu schaffen.

1.5. OSI-Schichtenmodell

Das OSI-Schichtenmodell (auch OSI-Referenzmodell oder OSI-Modell genannt) ist ein Schichtenmodell der Internationalen Organisation für Normung (ISO). Es wurde 1979 als Designgrundlage für das Zusammenspiel von Kommunikationsprotokollen in Rechnernetzen entwickelt und 1983 standardisiert. OSI steht hierbei für „Open Systems Interconnection" Reference Model. Das Schichtenmodell definiert sieben Schichten, jede Schicht stellt gewisse Funktionen zur Verfügung und jeder Schicht sind gewisse Protokolle zugeordnet. Ein Protokoll kann daher problemlos durch ein anderes Protokoll seiner Schicht ersetzt werden (z.B. TCP durch UDP -> beide sind aus der Schicht vier).

Nachfolgende schematische Darstellung zeigt den Datenfluss zwischen den sieben Schichten anhand von einer Verbindung zwischen Rechner A und Rechner B über einen Vermittlungsknoten (z.B. Router):

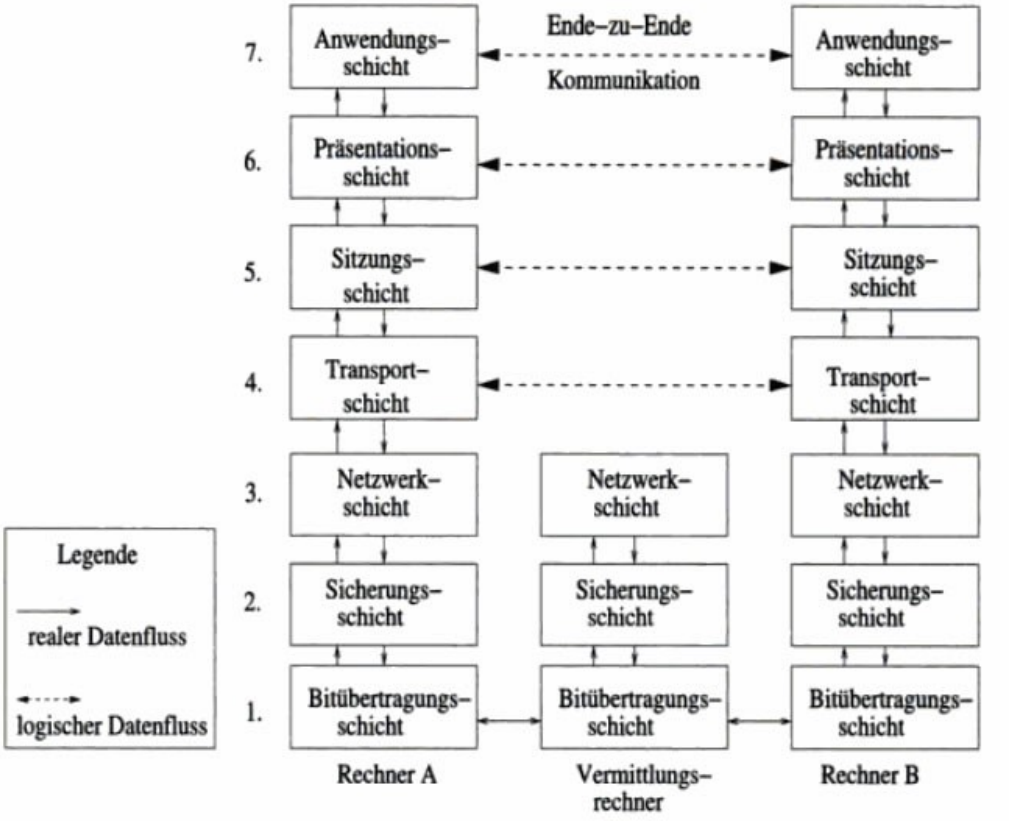

Abbildung 2: OSI-Schichtenmodell

[ECK06]

Kryptographie hat das Ziel den Inhalt einer Kommunikation zu schützen. Anonyme Kommunikation hingegen versucht die Beziehung zwischen den Kommunikationspartnern unkenntlich zu machen. Die Bereitstellung solcher Schutztechniken bei der Kommunikation in öffentlichen Netzwerken ist sehr herausfordernd und gewinnt zunehmend an Bedeutung. Im Fol-

genden werden die sieben Schichten des OSI Schichtenmodells näher definiert und erläutert. Es wird dabei auf die Anonymisierungsmöglichkeiten in der jeweiligen Schicht eingegangen:

1.5.1. Schicht 1, Bitübertragungsgsschicht (physical layer)

Die Schicht 1 ist für die physikalische Signalübertragung verantwortlich. Darunter fallen alle dafür erforderlichen Aufgaben wie zum Beispiel: Signalaufbereitung, Signalverstärkung, Signalwandlung, etc. Diese Schicht stellt jedoch keine Fehlerkorrektur oder Signalkontrolle zur Verfügung, dies ist Aufgabe der höheren Schichten. Des Weiteren ist auf dieser Ebene definiert wie überhaupt ein einzelnes Bit übertragen werden soll. Informationen werden als Bitfolgen gesehen, ein Bit kann dabei die Ausprägung 0 oder 1 haben. Für jedes Medium muss daher eine passende Codierung dieser Werte gefunden werden, beispielweise eine Funkwelle mit bestimmter Frequenz [TRV02].

Hardware auf dieser Schicht: Modem, Hub, Repeater, Schnittstellenvervielfacher, Stecker, etc.
Protokolle und Normen: V.24, V.28, X.21, RS 232, RS 422, RS 423, RS 499

Mögliche Anonymisierungstechniken:
Auf der Bitübertragungsschicht können nur wenige sinnvolle Anonymisierungsmaßnahmen getroffen werden. Der Grund dafür ist, dass die Geräte in der Lage sein müssen direkt miteinander zu kommunizieren. Beim allgemein vorherrschenden Ethernet Standard müssen die Geräte-Adressen kommuniziert werden, damit die Geräte miteinander kommunizieren können. Bei Bussystemen empfangen alle Stationen den von der sendeten Station gesendeten Inhalt. Es ist daher beispielsweise beim Token-Bus Verfahren nicht notwendig, dass eine Station die Adresse ihres Kommunikationspartners kennt um diesem Daten zu senden. Dies erleichtert die Anonymisierung erheblich. Theoretisch denkbar wäre es, die mögliche Kommunikation im Signal zu „verstecken“ (Steganographie) indem ein neuartiges Verfahren gewählt wird, welches nur die beiden Kommunikationspartner kennen. Da dies aber angepasste Hardware erfordern würde und kein Industriestandard verwendet werden könnte, wäre die mögliche Einsatzfähigkeit äußerst beschränkt.

1.5.2. Schicht 2, Sicherungsschicht (data link layer)

Aufgabe der Sicherungsschicht ist es, eine zuverlässige, d.h. möglichst fehlerfreie Datenübertragung sicher zu stellen und den Zugriff auf das Übertragungsmedium zu regeln. „Die Sicherungsschicht bündelt Bitströme der unteren Ebene zu Datenpaketen (engl. frames) bzw. teilt die Pakete der darüberlegenden Ebene auf Datenpakete auf und fügt Kontrollinformationen in Form von Prüfsummen hinzu, um Übertragungsfehler erkennen zu können.“ [ECK06] „Die Schicht 2 ist bei LANs nochmals unterteilt in die untere Teilschicht 2a (MAC-Schicht, media

access control), die den Zugriff auf das Übertragungsmedium inklusive Adressen (MAC-Adressen) regelt, und in die obere Teilschicht 2b (LLC-Schicht, logical link control), zu der die vom Medienzugriff unabhängigen Funktionen gehören.“ [TRV02]

Hardware auf dieser Schicht: Bridge, Layer-2-Switch
Protokolle und Normen: IEEE 802.11 (WLAN), IEEE 802.4 (Token Bus), IEEE 802.5 (Token Ring), FDDI

Mögliche Anonymisierungstechniken:
Mit einer Media Access Control (MAC) Adresse lassen sich Geräte im Netzwerk eindeutig identifizieren. Damit kann unter Umständen die jeweilige Seriennummer des Geräts herausgefunden werden. Mit dieser kann mit Hilfe externer Datenbanken ggf. der Käufer des Gerätes ermittelt werden. Dies lässt sich durch das Fälschen der MAC Adresse (MAC-Spoofing) verhindern. Diese Technik verwenden beispielsweise einige mobile Geräte in öffentlichen WLAN Netzwerken um das Erstellen von Bewegungsprofilen durch Dritte zu erschweren.

Mit einer MAC Adresse lassen sich evtl. auch die niederwertigen 64 Bits (der insgesamt 128 bit langen IPv6 Adresse) einer anonymisierten IPv6 Adresse rekonstruieren. Dieser Teil der Adresse wird auch Interface Identifier genannt und dieser wird im Normalfall mit Hilfe der eigenen MAC Adresse berechnet (siehe RFC4291 - Kapitel 2.5.1 „Interface Identifiers“). Diese Vorgehensweise lässt sich durch die Aktivierung der Privacy-Extensions (PEX) nach RFC 4941 verhindern, da durch die PEX-Aktivierung der Interface Identifier zufällig - anstatt auf der MAC Adresse basierend - generiert wird. Des Weiteren wechselt er periodisch.

1.5.3. Schicht 3, Vermittlungsschicht (network layer)

Bei leitungsorientierten Diensten sorgt die Vermittlungsschicht für das Schalten von Verbindungen und bei paketorientierten Diensten für die Weitervermittlung von Datenpaketen. Die Datenübertragung geht in beiden Fällen jeweils über das gesamte Kommunikationsnetz hinweg und schließt die Wegesuche (Routing) zwischen den einzelnen Netzknoten mit ein. D.h. im Gegensatz zu den darunter liegenden Schichten, die nur eine Verbindung mit benachbarten Knoten aufnehmen können, errichtet die Vermittlungsschicht eine Ende-zu-Ende Kommunikation zwischen den Kommunikationspartnern, die daher auch nicht direkt benachbart sein müssen. Weitervermittelte Pakete gelangen nicht in die höheren Schichten, sondern werden mit einem neuen Zwischenziel versehen und an den nächsten Knoten gesendet, so lange bis sie ihr Bestimmungsziel erreicht haben oder die Verbindung aus anderen Gründen abgebrochen wird.

Hardware auf dieser Schicht: Router, Layer-3-Switch
Protokolle und Normen: X.25, ISO 8208, ISO 8473 (CLNP), ISO 9542 (ESIS), IP, IPsec, ICMP

Mögliche Anonymisierungstechniken:

Mittels IP Adressen lassen sich Verbindungspartner eindeutig identifizieren, daher sind diese besonders im Fokus der Anonymisierung. IP Adressen können gefälscht werden (IP-Spoofing) dies hat aber die Einschränkung zur Folge, dass die Antwortpakete des Verbindungspartners an die gefälschte Adresse versandt werden und daher ein Zugriff auf diese in der Praxis äußerst schwer ist. Daher ist IP Spoofing meist nicht für die wechselseitige Kommunikation in geswitchten Netzwerken geeignet. IP Adressen lassen sich durch die Verwendung so genannter Anonymisierungsnetzwerke (TOR-Netzwerke) verschleiern. Diese Netzwerke schieben quasi Zwischenschritte in der Kommunikation ein, mit dem Ziel, dass der eigentliche Kommunikationspartner die ursprüngliche IP Adresse des Absenders nicht mehr erkennen kann. Diese Technik wird aufgrund des verwendeten Verschlüsselungsschemas auch als Onion-Routing bezeichnet. Nachteil dieser Art der Verschleierung sind Leistungseinbußen, da die Verbindung bewusst nicht auf dem kürzest möglichen Weg stattfindet und Ver- und Entschlüsselungsschritte notwendig sind. Eine einfachere Art die eigene IP Adresse zu verschleiern, stellt ein vorgeschalteter Proxy Server dar. In diesem Fall kennt zwar der Proxy Server die eigene IP Adresse, nicht jedoch die nachfolgenden Kommunikationspartner. Das gleiche Prinzip versuchen auch die verschiedenen Anonymous-Remailer Typen anzuwenden. Sei es durch Kombination verschiedener Anonymous-Remailer durch den Benutzer (WRAP-Prinzip) oder durch integrierte Verfahren wie das Mixmaster Konzept, welches eine Nachricht zuerst durch diverse Proxies leitet bevor diese den letzten Proxy (exit node) verlässt.

1.5.4. Schicht 4, Transportschicht (transport layer)

Hauptaufgabe der Transportschicht ist die Ende-zu-Ende-Fehlerkontrolle. Also die Sicherstellung der korrekten Datenübertragung zwischen den Kommunikationspartnern. Des Weiteren gehören noch verschiedene andere Dienste wie z.B. die Segmentierung von Datenpaketen und die Stauvermeidung (engl. congestion avoidance) zu dieser Schicht. Die Transportschicht bietet den anwendungsorientierten Schichten 5 bis 7 einen einheitlichen Zugriff, so dass diese die Eigenschaften des Kommunikationsnetzes nicht zu berücksichtigen brauchen.

Hardware auf dieser Schicht: Gateway, Content-Switch
Protokolle und Normen: ISO 8073/X.224, ISO 8602, TCP, UDP, SCTP.

Mögliche Anonymisierungstechniken:

Virtual Private Networks übertragen Daten innerhalb eines kryptographischen Tunnels, der zwischen Benutzer und VPN-Anbieter etabliert wird, um Vertraulichkeit, Authentifizierung und Integrität der übertragenen Kommunikation zu gewährleisten. Typische VPN-Technologien arbeiten mindestens auf der Transportschicht und sind in der Lage beliebigen Netzwerkverkehr zu kapseln. Daher können sie vertrauliche Daten zum Ziel senden ohne, dass die Gefahr besteht, dass am Zwischenweg diese vertraulichen Daten eingesehen werden können. Beispielsweise werden öffentliche VPN-Anbieter im Internet dazu verwendet den eigenen Standort zu anonymisieren um so genannte Ländersperren von Inhalten (z.B. Musik, Filme) zu umgehen.

Jede TCP Verbindung, die ein System verwendet wird eindeutig durch die folgenden vier Eigenschaften identifiziert: die lokale Adresse, der lokale Port, die remote Adresse und der remote Port. Der lokale Port für eine ausgehende Verbindung wird vom Betriebssystem selbstständig ausgewählt und nach dem Ende der Verbindung wieder freigegeben. Da dieser Vorgang für den Benutzer transparent erfolgt wird der gewählte Port auch anonymer Port genannt. Durch den Aufbau mehrerer, gleichzeitiger Verbindungen zu einem remote Computer, die sich nur durch den lokalen Port unterscheiden, lassen sich beispielsweise mehre Benutzer (dies wird bei Lasttests angewandt) simulieren.

1.5.5. Schicht 5, Sitzungsschicht (session layer)

Die Sitzungsschicht koordiniert und synchronisiert die Kommunikation zwischen Anwendungsprozessen und trifft Vorkehrungen für ein Wiederaufsetzen unterbrochener Sitzungen. Dazu integriert sie Informationen über Sicherungspunkte (checkpoints) in den Nachrichtenstrom. Anmerkung: Funktionalitäten der Schicht 5 werden meistens von Protokollen der Schicht 7 mit implementiert daher hat diese Schicht als eigenständige Schicht wenig Bedeutung erlangt.

Hardware auf dieser Schicht: Gateway, Content-Switch
Protokolle und Normen: ISO 8306 / X.215 (Session Service), ISO 8327 / X.225 (Connection-Oriented Session Protocol), ISO 9548 (Connectionless Session Protocol)

Mögliche Anonymisierungstechniken:

Secure Socket Layer (SSL) und Transport Layer Security (TLS) sind zwei äußerst bekannte Protokolle der Sitzungsschicht. Mit ihnen lässt sich Internet Kommunikation zwischen Kommunikationspartnern gegenüber Dritten absichern. Aufgabe der Sitzungsschicht ist es die Kommunikation in eine eindeutige Sitzung (Session) einzuteilen. Die Programmiersprache Java ermöglicht es beispielsweise Informationen über den Benutzer in so genannten „session beans“ abzuspeichern, die so lange gespeichert werden, wie die Sitzung aufrecht bleibt.

1.5.6. Schicht 6, Präsentationsschicht (presentation layer)

„Da in einem Rechnernetz unterschiedliche Datenformate existieren, übernimmt die Darstellungsschicht die Aufgabe, die Daten der Anwendungsschicht, also die maschinenabhängigen Repräsentationen, in ein netzeinheitliches Format umzuwandeln. Im Gegensatz zu den unteren Schichten, die vordringlich reine Transportaufgaben durchführen ohne die Nachrichteninhalte zu interpretieren, beschäftigt sich diese Schicht mit der Syntax und Semantik der übertragenden Informationen. Zur Beschreibung der hierzu eingesetzten Transfersyntax werden Sprachen definiert wie ASN (Abstract Notation No 1). Weitere Aufgaben der Informationsrepräsentation betreffen die Datenkompression oder die Datenverschlüsselung, die häufig auf dieser Schicht angesiedelt werden.“ [ECK06] Anmerkung: Funktionalitäten der Schicht 6 werden meistens von Protokollen der Schicht 7 mit implementiert daher hat diese Schicht als eigenständige Schicht wenig Bedeutung erlangt.

Hardware auf dieser Schicht: Gateway, Content-Switch
Protokolle und Normen: ISO 8822 / X.216 (Presentation Service), ISO 8823 / X.226 (Connection-Oriented Presentation Protocol), ISO 9576 (Connectionless Presentation Protocol)

Mögliche Anonymisierungstechniken:
Es ist keine gesonderte Bewertung für die Präsentationsschicht möglich, da diese Schicht sehr eng mit der Anwendungsschicht verknüpft ist.

1.5.7. Schicht 7, Anwendungsschicht (application layer)

Diese (oberste) Schicht verschafft den Anwendungen Zugriff auf das Netz. Beispielsweise der Austausch und das Verschieben von Daten über das Netzwerk, Rechneraufträge an entfernte Rechner, das Versenden von E-Mails etc. Die Anwendungen selbst gehören nicht zu dieser Schicht. Zu dieser Schicht lässt sich eine Vielzahl an Protokollen rechnen.

Bereits die Transportschicht leistet viel Vermittlungsarbeit zwischen Applikationen und dem Netzwerk. Dennoch ist die Anwendungsschicht essentiell, da sie speziell auf die Netzwerk Dienste (z.B. Domain-Name-System), APIs (Application-Programming-Interfaces, z.B. NetBIOS), unterschiedliche Betriebssysteme und weitere spezielle Gegebenheiten der Applikationen bzw. Applikationsumgebung eingeht. Durch die Anwendungsschicht werden Applikationen entkoppelt und müssen sich nicht um die Besonderheiten des jeweiligen Netzwerks kümmern. Des Weiteren werden von den jeweiligen Anwendungsprotokollen standardisierte APIs bereitgestellt um den Applikationen einen einfachen Zugriff auf Netzwerk Funktionalitäten zu ermöglichen.

Hardware auf dieser Schicht: Gateway, Content-Switch, Protokollumwandler
Protokolle und Normen: X.400, X.500, ISO 8571 (FTAM), ISO 9040/9041 (VT), ISO 9506 (MMS), MHS, VTP, FTP, NFS, SMTP, POP3, IMAP, HTTP, LDAP, JTM, SSH, etc.

Mögliche Anonymisierungstechniken:

Gewonnene Informationen aus der Anwendungsschicht können dazu dienen, ausführliche Personenprofile zu erstellen. Wenn eine Anwendung beispielsweise persönliche Daten überträgt, ist die Anonymisierung darunter liegender Schichten hinfällig. Auf den ersten Blick unverdächtige Daten aus der Anwendungsschicht können in Kombination mit Daten aus der Vermittlungsschicht die De-Anonymisierung deutlich beschleunigen. Beispiele dafür wären der verwendete Browser, die verwendete Browserversion, die Betriebssystemversion, Cookies und Anmeldedaten. Diese Daten werden im Normalfall von jedem Browser automatisch übermittelt. Alle eindeutigen Eigenschaften, die es einem Angreifer ermöglichen kommunizierende Entitäten voneinander abzugrenzen müssen möglichst vermieden werden.
Für den Anonymous Remailer sind die Metadaten, die von E-Mail Protokollen hinzugefügt werden, von besonderer Relevanz. Aus den Kopfdaten, die oft für den Benutzer transparent sind, lassen sich genügend Informationen extrahieren um den ursprünglichen Absender zu ermitteln. Daher wird in dieser Masterarbeit ausführlich auf die Metadaten im E-Mail Verkehr eingegangen und anhand des Anonymous Remailers wird erklärt was beachtet werden muss um eine anonyme Kommunikation zu ermöglichen.

1.6. TCP/IP-Schichtenmodell

Um Probleme der Netzwerkkommunikation im Allgemeinen zu betrachten, greift man auf das im vorherigen Kapitel beschriebene ISO/OSI-Referenzmodell zurück. Das TCP/IP-Referenzmodell hingegen ist speziell auf die Internet-Protokolle zugeschnitten, die den Datenaustausch über die Grenzen lokaler Netzwerke hinaus ermöglichen. Es wird weder der Zugriff auf ein Übertragungsmedium noch die Datenübertragungstechnik definiert. Vielmehr sind die Internet-Protokolle dafür zuständig, Datenpakete über mehrere Punkt-zu-Punkt-Verbindungen (Hops) weiterzuvermitteln und auf dieser Basis Verbindungen zwischen Netzwerkteilnehmern über mehrere Hops herzustellen.

Applikationsprotokolle

Der größte Unterschied zum ISO/OSI-Referenzmodell besteht darin, dass das TCP/IP-Referenzmodell nur 4 Schichten definiert. Im konkreten werden die Schichten 1 und 2 sowie die Schichten 5 bis 7 zu jeweils einer Schicht zusammengefasst. Die Schichten 5 bis 7 werden im TCP/IP Modell als Anwendungsschicht oder auch Application Layer bezeichnet. Bei Internetprotokollen ist es zweckmäßig die TCP/IP Definition der Anwendungsschicht zu verwenden, da z.B. viele Protokolle der Schicht 7, gleichzeitig auch die Schichten 5 und 6 mit

implementieren. Die meisten Applikationsprotokolle sind u.a. für die Kommunikation über das Internet ausgelegt.

2. Aufgabenstellung

Ziel der Masterarbeit ist es einen Cypherpunk Anonymous Remailer (Type I) mit besonderem Fokus auf Natural-Language-Processing zu entwickeln. Dieser muss u.a. folgende Anforderungen erfüllen.

2.1. E-Mail Header Anonymisierung

Der Remailer muss mittels SMTP E-Mail Nachrichten empfangen können und deren E-Mail Header anonymisieren. Es müssen dabei alle Header entfernt werden, die einen Rückschluss auf den eigentlichen Sender ermöglichen. Eine umfangreiche Liste bestehender E-Mail Header findet man im RFC 2076 vor: http://www.ietf.org/rfc/rfc2076.txt

Die in der folgenden Auflistung bekannter E-Mail Header **fett** gedruckten Header Elemente werden für den korrekten Remailer Betrieb benötigt. Sie dürfen daher nicht entfernt werden, müssen jedoch sofern notwendig anonymisiert werden. Die restlichen Header Einträge werden entfernt, da sie für den E-Mail Versand nicht notwendig sind und dem Empfänger Informationen über den ursprünglichen Absender verraten könnten. Beim Weiterleiten des Emails werden die für den E-Mail Versand notwendigen Header (z.B. MAIL FROM, RCPT TO, FROM, TO) neu erstellt. Diese Header werden Envelope-Header genannt. Es folgt eine (nicht vollständige) Auflistung bekannter E-Mail Header inklusive deren Bedeutung.

- Envelope-From: Parameter MAIL FROM von SMTP Handshake
- Envelope-To: Parameter RCPT TO von SMTP Handshake
- From: Absender (der verantwortliche Absender)
- To: Der Empfänger
- CC (optional): Carbon Copy
- BCC (optional): Blind Carbon Copy
- **Subject: Betreffzeile**
- Date: Absendedatum und Uhrzeit (Lokalzeit + Abweichungsangabe von UTC)
- **MIME-Version: Angabe der verwendeten MIME Versionsnummer**
- **Content-Type : Art der MIME Codierung des Mail Inhalts und die Kennung der jeweiligen Teil-Abgrenzungen (boundary)**
- Delivery-Date (optional): Empfangsdatum der E-Mail
- Received: Der bisherige Weg, den das E-Mail genommen hat (mehrere Received: Einträge)

- Return-Path (optional): Rücksenden an, falls unzustellbar, (anhand von MAIL FROM nachgetragen)
- References (optional): Referenz auf eine andere E-Mail, Message-ID dieser anderen Nachricht
- Authentication-Results (optional): Senderauth. Information, für SPF (Sender Policy Framework)
- DKIM-Signature (optional): DomainKeys ist ein Identifikationsprotokoll zur Sicherstellung der Authentizität von E-Mail-Absendern
- Sender (optional): Technischer Absender (z.B. bei einer Mailingliste)
- Organization (optional): Die Organisation des Absenders, falls zutreffend
- Reply-To (optional): Antwortadresse(n)
- Delivered-To (Optional): Empfänger (wie To:), wird evtl. vom Mailhoster des Empfängers eingefügt
- Message-Id (optional): Eindeutige Nachrichten ID
- Disposition-Notification-To (optional): RFC 2298, der Absender fordert eine (Lese-) Bestätigung
- X-Mailer (optional): Name und Versionsnummer des zum Versenden genutzten Mailprogramms
- X-Mailer und X-MSMail-Priority (optional): Prioritätsangaben (nur für Sender und Empfänger relevant)
- X-Sender (optional): Hier wird der Wert von MAIL FROM eingetragen.
- -spam und virenscanner tags z.B. x-scanned-by, alle weiteren X-* tags

Das folgende Beispiel veranschaulicht welche Header auf den Absender schließen lassen (rote Schrift), welche Header auf dem Empfänger schließen lassen (blaue Schrift) und welche Header sich neutral verhalten (schwarze Schrift):

```
Return-Path: <donotreply@register.jajah.com>
Delivered-To: GMX delivery to soellnerrobert@gmx.at
Received: (qmail invoked by alias); 18 Jun 2012 11:39:36 -0000
Received:   from   mailservice3.jajah.com   (EHLO   mailservice3.jajah.com)
[91.194.4.145]
  by mx0.gmx.net (mx073) with SMTP; 18 Jun 2012 13:39:36 +0200
Received: from WEB46 (web46.jajah.dublin [192.168.52.16])
      by mailservice3.jajah.com (Postfix) with ESMTP id D201133986A
      for <soellnerrobert@gmx.at>; Mon, 18 Jun 2012 11:39:34 +0000 (UTC)
MIME-Version: 1.0
From: billing@jajah.com
Sender: donotreply@register.jajah.com
To: soellnerrobert@gmx.at
Reply-To: support@jajah.com
Date: 18 Jun 2012 12:39:34 +0100
Subject: =?utf-8?B?SkFKQUggWmFobHVuZ3NiZXN0w6R0aWd1bmc=?=
Content-Type: multipart/mixed;
```

```
 boundary=--boundary_12591_0747beee-0c92-47d2-b512-ad8326cf07b0
Message-Id: <20120618113934.D201133986A@mailservice3.jajah.com>
X-GMX-Antivirus: 0 (no virus found)
X-GMX-Antispam: 0 (Sender is in whitelist: billing@jajah.com);
 De-
tail=5D7Q89H36p4L00VTXC6D4q0N+AH0PUCnD09sT3te/pmSYe4NcswD1SSFCNqH2gzeZvZZg
 DomvTwx3KO7GGkDhA==V1;
X-AntiVirus: checked (incoming) by Avira MailGuard (Version: 12.3.1.15;
AVE:8.2.10.92; VDF:7.11.33.72
```

2.2. Weitere Hauptanforderungen

Es folgt eine Auflistung geforderter zusätzlicher Anforderungen für den Anonymous Remailer:

- Der Remailer soll eingehende Nachrichten vor dem Weitersenden zufällig verzögern (Delay time). Dies inkludiert eine mögliche Umsortierung der eingehenden Nachrichten zwischen Nachrichtenempfang und Weiterleitung. Ziel dieser Anforderung ist es, Beobachtern die Zuordnung von eingehenden zu ausgehenden Nachrichten zu erschweren, da die zeitliche Zusammengehörigkeit gestört wird.

Um diese Maßnahme noch effektiver zu gestalten, sollen auch zwei Anforderungen zur Veränderung der Größe, der ausgehenden (anonymisierten) E-Mails erfolgen. Dies soll es Angreifern erschweren anhand eines Größenvergleichs zwischen eingehender und ausgehender Nachricht die beiden Nachrichten einander zuordnen zu können:

- Es soll eine zufällige Anzahl an künstlich erzeugten Header Elementen einer Nachricht hinzugefügt werden. Die Header Elemente sollen selbst spezifiziert werden können. Z.B. X-Header z.B. „X-Company" oder „X-Location“ mit diversen vordefinierten Werten.
- Es soll ein zufällig erzeugter Text an den Nachrichtenbody angehängt werden um die Nachrichtengröße zu verändern. Beispielsweise ein „lorem ipsum“ text.

Es sollen sowohl der eigentliche Nachrichteninhalt als auch Attachements (Bilder, PDFs, etc.) einer Überprüfung unterzogen werden:

- Der Nachrichtentext (body) und der Betreff (Subject) sollen gescannt werden. Mit Hilfe einer Technik zur Namenserkennung/Filterung dem Natural Language Processing (NLP), sollten Personennamen, Orte, Organisationen erkannt und entfernt werden. Der entsprechende Platzhalter für das Ersetzen soll frei wählbar sein.
- Der Remailer soll Attachements, die mit den eingehenden Nachrichten mitversandt wurden, entfernen. Da deren Inhalt nicht überprüft werden kann (z.B. Bilder) bzw. eine Prüfung jedes Attachments zu zeitaufwendig wäre, soll im Nachrichtentext ein Verweis eingefügt werden, dass das entsprechende Attachment entfernt wurde.
- Der Remailer soll über ein graphisches Interface (GUI) bedienbar sein, indem die notwendigen Parameter für den Remailerbetrieb und die Anonymisierung eingegeben werden. Das sind u.a.:

- Alle für den SMTP Serverbetrieb (eingehend und ausgehend) notwendigen Parameter
- Der minimale (minDelay) und maximale (maxDelay) Nachrichten-Verzögerungswert (Eingabe in Sekunden)
- Option ob nach Namen und/oder Orten und/oder Organisationen gescannt werden soll (damit diese ersetzt werden). Texteingabefeld, wo der fix hinterlegte Name geändert werden kann, mit dem im Mail gefundene Namen, Orte und Organisationen ersetzt werden.

3. Verwendete Techniken und Protokolle

Dieses Kapitel beschreibt die für den E-Mail Versand und Empfang notwendigen Protokolle und Techniken. POP3, IMAP und SMTP werden kurz erklärt und es wird für jedes Protokoll auf die Anonymisierungsmöglichkeiten eingegangen. Zusätzlich wird bei jedem Protokoll auf die sicherheitstechnisch relevanten Punkte eingegangen. Des Weiteren wird eine für die E-Mail Kommunikation äußerst wichtige Technik vorgestellt. Die (Secure) Multipurpose Internet Mail Extensions (S/MIME). Ohne diese Technik wäre die E-Mail Kommunikation – in der jetzigen Form – nicht möglich.

3.1. (Secure) Multipurpose Internet Mail Extensions (S/MIME)

Aktuelle RFC Standards: RFC 2045 bis RFC 2049 (alle 1996), RFC 4288, RFC 4289 (2005)

Multipurpose Internet Mail Extensions (MIME) ist ein Standard, der die Struktur und den Aufbau von E-Mails und anderen Internetnachrichten festlegt. MIME ermöglicht es E-Mails zu senden bzw. zu empfangen, die beispielsweise:

- eine andere Textkodierung verwenden als ASCII.
- ein Attachement haben.
- mehrere unterschiedliche Attachements in nur einer E-Mail Nachricht kodieren.
- Header Informationen verwenden, die nicht mit ASCII kodiert sind.

Ferner findet MIME Anwendung bei der Deklaration von Inhalten in verschiedenen Internetprotokollen, so zum Beispiel in HTTP und bei Desktop-Umgebungen z.B. KDE, GNOME, XFCE oder Aqua.

MIME wurde entwickelt um einige ernstzunehmende Einschränkungen beim Senden/Empfangen von E-Mail aufzuheben. Die größte Einschränkung war, dass das Basis E-Mail Versand Protokoll SMTP nur die 7-bit ASCII Textkodierung unterstützt. Daher war nur der Versand von Nachrichten in einer stark limitierten Anzahl von Sprachen möglich, hauptsächlich in Englisch. Der MIME Standard wurde 1992 in zwei RFCs (1341 und 1342) veröffentlicht. Ein Jahr später wurden diese durch die RFCs 1521 und 1522 ersetzt. 1994 wurde mit dem RFC 1590 ein ergänzender RFC veröffentlicht der dokumentiert wie neue MIME Typen definiert werden sollen. Im November 1996 wurde der Standard erneut überarbeitet und es wurden 5 neue RFCs (2045 bis 2049) veröffentlich die den MIME Standard in fünf Teilen beschreiben. Dieser Schritt sollte vor allem die Lesbarkeit der Informationen verbessern. Im Laufe der Zeit kamen noch etliche, ergänzende RFCs hinzu. [SCW10]

Funktionsweise

Im RFC 822 wurde festgelegt, dass E-Mail Nachrichten nur aus ASCII-Text bestehen dürfen. MIME umgeht diese Beschränkung elegant indem einfach Nicht-ASCII-Text (dies inkludiert auch alle Arten von Attachements wie z.B. Multimedia Dateien), mittels ASCII-Text kodiert wird. Um dies zu erreichen wurde ein umfassendes Kodierungssystem entworfen. Damit der Empfänger einer MIME kodierten Nachricht dennoch weiß, wie die Nachricht zu dekodieren ist, wurden neue Einträge dem Header der Nachricht hinzugefügt. Da nun der ganze Text in ASCII kodiert vorliegt, ist die Forderung von RFC 822 erfüllt. So mussten die anderen E-Mail Protokolle nicht verändert werden, nur Sender und Empfänger müssen jeweils MIME kompatible E-Mail Programme verwenden um eine Nachricht senden/empfangen zu können (heutzutage sind alle verwendeten E-Mail Programme MIME kompatibel). [KOZ11]

Dieser „Trick" war möglich da der RFC 822 zwar die Text Kodierung auf ASCII beschränkt und vorsieht, dass jede Zeile nur maximal 998 Zeichen lang ist sowie mit „CRLF" abgeschlossen werden muss, aber der Inhalt der Nachricht kann frei gewählt werden d.h. insbesondere der Text kann auch nur maschinenlesbar sein. Des Weiteren definierte RFC 822 sogenannte „user-defined headers", also Header Einträge die benutzerspezifisch einer Nachricht hinzugefügt werden können. Wie die Kodierung der Nachricht und die verwendeten MIME Header genau aussehen hängt von der Art der MIME Nachricht ab. Prinzipiell gibt es zwei verschiedene Nachrichtentypen: [KOZ11]

- Simple (Discrete Media) Structure
  Wird verwendet für Nachrichten die nur einen Content Type verwenden z.B. nur Text oder nur ein Bild.
- Complex (Composite Media) Structure
  Wird verwendet für Nachrichten die mehrere Kodierungen im Nachrichten Body verwenden z.B. eine Nachricht mit angehängtem Bild oder mehre Attachements. Content-Type immer „multipart/*".

Die wichtigsten fünf MIME Header Einträge (definiert in RFC 2045):

Header	**Bedeutung**
MIME-Version	Zeigt an, dass die Nachricht mit MIME kodiert wurde sowie die verwendete Version (bisher nur Version 1.0)
Content-Type	Gibt den Typ der MIME kodierten Daten an. Besteht immer aus einem Content Type und einem Subtype, getrennt durch einen Schrägstrich. Im Header sagt dieser Eintrag aus ob ein Simple type z.B. text/html oder ein Complex type verwendet wird z.B. multipart/mixed Im Fall eines Complex Type kommt dieser Eintrag auch im Nachrichten Body vor und gibt dort jeweils pro verwendeten Typ, die Art des Typs an.

Boundary	Dieser Tag kommt in Kombination mit dem Content-Type: multipart vor. Es wird jene Zeichenfolge angegeben, die die einzelnen Abschnitte einer multipart-Nachricht voneinander abgrenzt. Die boundary ist ein generierter Wert aus Zeichen und Zahlen. Es wird ein Abschnittswechsel erkannt, wenn der im Header definierte boundary Wert mit dem Prefix „--“ im Body vorkommt. Es ist sicherzustellen, dass der definierte boundary Wert mit dem Prefix „--“ nicht im normalen Body Text vorkommt, da sonst fälschlicherweise ein neuer Abschnitt interpretiert werden würde.
Content-Transfer-Encoding	Gibt die genaue Kodierungsart pro verwendeten Typ an. Der Default Wert ist 7bit, welches der gewöhnlichen Kodierung von ASCII-Text entspricht. Oft verwendet wird 8bit im Zusammenspiel mit UTF-8, weitere Kodierungsarten: binary, quoted-printable, base64, …
Content-ID	Erlaubt es MIME Typen eine bestimmte ID zuzuordnen. Ist optional und wird meistens bei Multpart Nachrichten im Body eingesetzt.
Content-Description	Optional, um eine textuelle Beschreibung für einen MIME Typ zu definieren.

Sicherheitsüberlegungen

Die als S/MIME bekannte Erweiterung ermöglicht unter anderem die verschlüsselte und digital signierte Übertragung von E-Mail Nachrichten. Dazu wurden zwei zusätzliche Contenttypes definiert: Das Multipart/Signed-Format um E-Mail Nachrichten digital zu signieren (Signatur wird über MIME Header und Nachrichten Content gebildet und als Block an die Nachricht angehängt. Dadurch bleibt diese auch für Clients lesbar, die das S/MIME Format nicht unterstützen) und das Multipart/Encrypted-Format um diese zu verschlüsseln (Im ersten Block sind die Informationen zur Entschlüsselung vorhanden, im zweiten Block die verschlüsselten Daten. Spam- und Virenscan kann daher erst beim Empfänger erfolgen, der Betreff wird allerdings nicht verschlüsselt, daher keine geheimen Informationen in den Betreff!). S/MIME wird von den meisten modernen Mailclients unterstützt. Es erfordert X.509-basierte Zertifikate für den Betrieb. [KOZ11]

3.1.1. MIME Anonymisierungsüberlegungen

MIME ist für die Entwicklung des Anonymous Remailer eine äußerst wichtige Technik, da E-Mails viele im MIME-Format codierte Informationen enthalten. Diese Informationen können zur Identifizierung des Absenders verwendet werden, daher ist eine sorgfältige Anonymisierung der Daten erforderlich. Es ist zu beachten, dass die Boundary Angabe eine E-Mail Nachricht eindeutig kennzeichnet und dadurch eine Rückverfolgbarkeit gegeben ist, falls die Boundary Angabe bei der Weiterleitung von E-Mail Nachrichten nicht geändert wird. Daher sendet der Anonymous Remailer E-Mails prinzipiell im MIME Format Content-Type: multipart/mixed; mit einer neu generierten Boundary weiter, um Rückschlüsse zwischen eingegan-

gen und gesendeten E-Mails anhand der zwangsläufig enthaltenen MIME Klassifikation zu verhindern.

3.2. Post Office Protocol Version 3 (POP3)

Protokoll:	**Post Office Protocol Version 3 (POP3)**
Standards (aktuell):	RFC 1939 (1996)
Standardport(s):	TCP Port 110 und TCP Port 995 (verschlüsselt)
Verbreitung:	Sehr Hoch
Schadenspotenzial:	Sehr Hoch
Funktionsgruppe:	E-Mail Protokolle
Zweck:	POP3 ist das letzte Glied in der Kette des E-Mail Versands, mittels diesem Protokoll kann ein Empfänger seine E-Mail Nachrichten vom seinem SMTP-Server abrufen. Dazu ist keine permanente Verbindung nötig, der SMTP-Server speichert empfangene Nachrichten vorläufig ab und bei der nächsten Verbindung des Clients zum Server werden die Nachrichten dann an den Client übertragen. POP3 ist dabei ein relativ simples, zustandsbasiertes Protokoll und unterstützt zusammengefasst nur das Auflisten, Abholen und Löschen von E-Mails am SMTP-Server.

Funktionsweise

Das Designziel des Post Office Protocols war der Grundsatz: Keep it simple! So war beispielsweise der RFC 918 (1984) - die erste Version von POP - nur 5 (!) Seiten lang. Aber auch der erste RFC des POP Protokolls in der Version 3, der 1988 veröffentlichte RFC 1081, kam mit ‚nur' 15 Seiten aus. Der aktuelle RFC 1939 umfasst 22 Seiten und ist damit, im Vergleich zu anderen Protokollbeschreibungen, immer noch sehr kompakt.

POP3 arbeitet zustandsbasiert. Ein POP3-Server wartet auf TCP Port 110 (bzw. Port 995, falls verschlüsselt) auf eingehende Befehle des POP3-(Mail)Clients. POP3 verwendet ebenso wie FTP oder SMTP textuelle Befehle (typischerweise 3 bis 4 Zeichen lang), im Gegensatz zu diesen Protokollen werden aber die komplexen dreistelligen Replycodes für die Antworten nach jedem Befehl nicht verwendet sondern nur folgende Unterscheidung:

- +OK Befehl erfolgreich durchgeführt
- -ERR Es ist ein Fehler aufgetreten, es folgt meistens eine kurze Fehlerbeschreibung

POP3 lässt sich auch als deterministischer Zustandsautomat beschreiben, da es immer dieselben Phasen, nacheinander durchläuft:

1. **Verbindungsaufbau:** Der Client baut die TCP Verbindung zum Server auf.
2. **Authorization State:** Der Server signalisiert dem Client, dass er bereit ist Befehle zu empfangen, der Client authentifiziert sich anschließend am Server (z.B. durch die Befehle „USER“ und „PASS“, deren Übermittlung ungeschützt, im Klartext erfolgt) um Zugang zu seiner Mailbox zu bekommen.
3. **Transaction State:** Der Client darf dem Server Befehle schicken wie z.B. das Anzeigen, Herunterladen oder Löschen von Nachrichten (setzt ein Flag).
4. **Update State:** Der Client beendet die Transaction Phase durch das Senden von „QUIT“, dadurch wechselt der Server automatisch in den Update State und führt diverse CleanUp Befehle durch z.B. das Löschen von Nachrichten, bei denen das Deletion Flag gesetzt wurde.
5. **Verbindungsabbau:** Sind alle Aktionen abgeschlossen wird die Sitzung beendet und die TCP Verbindung abgebaut.

Sicherheitsüberlegungen

Neben dem einfachen USER/PASS Mechanismus gibt es auch noch andere Authentifizierungsarten, die die Sicherheit erhöhen. Beispielsweise der APOP Befehl, der mittels eines eindeutigen Zeitstempels und dem Passwort einen MD5 Hashwert berechnet und diesen an den Server überträgt. Eine andere Möglichkeit ist die Verwendung alternativer Authentifizierungsmethoden durch den AUTH Befehl, wie in RFC 1734 beschrieben. Die Übertragung der Authentifizierungsdaten, der POP3-Kommandos und der Nachricht selbst kann auch komplett über SSL/TLS verschlüsselt erfolgen. Dabei wird alternativ das STARTTLS-Verfahren (das Kommando lautet STLS) auf dem Standard-TCP-Port 110 benutzt, wobei hier der Client die Verschlüsselung erzwingen muss, da sonst die Verschlüsselung – für ihn unbemerkt – deaktiviert werden kann. Die gebräuchlichste Variante stellt POP3 über SSL (POP3S) dar, die Verbindung läuft in diesem Fall über TCP-Port 995. Für nähere Information siehe RFC 2595 – „Using TLS with IMAP, POP3 and ACAP“.

3.2.1. POP3 Anonymisierungsüberlegungen

Eine Anonymisierung der eigenen Verbindungsdaten ist bei POP3 schwer zu erreichen. Einerseits sind normalerweise bei der Anmeldung auf dem POP3 Server ein Username und das zugehörige Passwort einzugeben. Per Default findet dies im Klartext statt. Dadurch kann der Server mehrere Mailkonten auf seinem Server verwalten. Aus Anonymisierungssicht kann jedoch auch nachvollzogen werden, wann, welcher Benutzer sich mit dem Mailserver zu welchem Zeitpunkt verbunden hat. Andererseits wird auch die IP Adresse des Benutzers, der sich mit dem Server verbindet, an den Server übermittelt. Dies kann die Zuordnung des Mailkontos zu einem Individuum ermöglichen (siehe Abschnitt 1.5

OSI-Schichtenmodell). Um dies zu verhindern ist ein zweistufiges Schutzkonzept erforderlich. Um Dritten das Abhören der Verbindung zu erschweren, sollte die verschlüsselte Protokollvariante POP3S verwendet werden (siehe obigen Abschnitt Sicherheitsüberlegungen). Möchte man auch vermeiden, dass der Server die tatsächliche IP Adresse erhält, so muss der Benutzer ein Anonymisierungsnetzwerk verwenden z.B. TOR, um seine eigene IP Adresse zu verschleiern.

3.3. Internet Message Access Protocol Version 4 (IMAP4)

Protokoll:	**Internet Message Access Protocol Version 4 (IMAP4)**
Standards (aktuell):	RFC 3501 (IMAPv4, 2003), zahlreiche opt. Protokollerweiterungen
Standardport(s):	TCP Port 143 oder TCP Port 993 (verschlüsselt)
Verbreitung:	Hoch
Schadenspotenzial:	Sehr Hoch
Funktionsgruppe:	E-Mail Protokolle
Zweck:	IMAP erfüllt prinzipiell denselben Zweck wie POP3 und ermöglicht es Empfängern von E-Mail Nachrichten, diese vom jeweils zuständigen SMTP-Server abzurufen. IMAP bietet jedoch einen deutlich größeren Funktionsumfang als POP3 und zahlreiche Vorteile, jedoch mit dem Nachteil der wesentlich größeren Komplexität und damit verbundenen aufwendigen Implementierbarkeit. Der größte Vorteil von IMAP ist, dass die E-Mails immer am Server verbleiben und damit sowohl die Verwaltung (auch bei mehreren Clients) als auch die Sicherung der Daten zentral erfolgen können.

Das Internet Message Access Protocol kann auf eine durchaus komplizierte Vergangenheit zurück blicken. Es wurde 1986 mit dem Ziel entwickelt den Zugriff auf Mailboxen und Nachrichten so bereitzustellen, als befänden sich diese direkt auf dem lokalen Rechner. Die erste Protokollversion, die als Internetstandard aufschien, war die IMAP Version 2, die 1988 im RFC 1064 veröffentlicht wurde. Danach folgte Version 3, die aber nie wirklich akzeptiert wurde und ein Gegenvorschlag dazu, der als „IMAP2bis“ bekannt wurde, sich aber ebenfalls nicht durchsetzen konnte. Die heutige Version 4 wurde 1994 in den RFCs 1730 (Core) und 1731 (Authentication) veröffentlich. Diese RFCs wurden danach noch weiter verbessert, bis schließlich 2003 die aktuelle Version IMAP4rev1 im RFC 3501 erschien. Diese Protokollversion ist heutzutage fast ausschließlich in Verwendung.

Funktionsweise

IMAP arbeitet genau wie POP3 zustands- und textbasiert, Die folgende Abbildung illustriert die möglichen Zustände bei IMAP in einem Zustandsdiagramm: [KOI11]

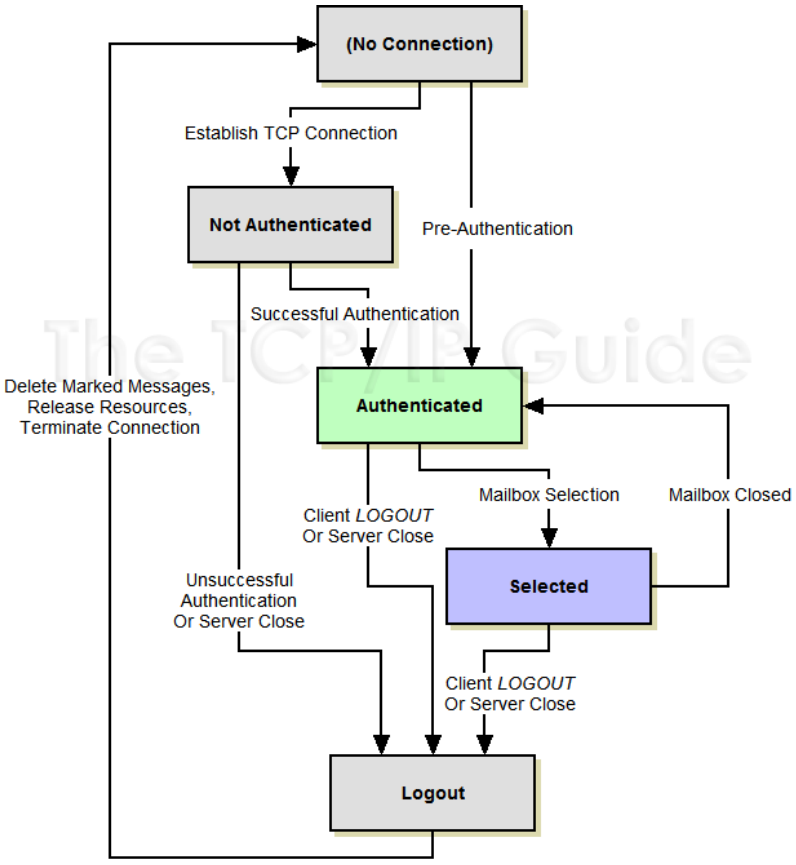

Abbildung 3: IMAP Zustandsdiagramm

IMAP definiert einen Befehlssatz der, im Gegensatz zu den meisten anderen Protokollen, nicht abgekürzt wird. So wird z.B. STATUS nicht wie in POP3 als STAT sondern wirklich als STATUS übertragen. Der aktuell anwendbare Befehlssatz hängt vom jeweiligen Zustand ab, indem sich die Verbindung gerade befindet. Jeder Befehl vom Client an den Server enthält außerdem einen eindeutigen ‚Tag', dies ist eine alphanummerische Nummer die sich inkrementell erhöht. Dieses Konzept ermöglicht es dem Server auf jeweils einen bestimmten Befehl des Clients zu antworten, wenn dieser mehrere gesendet hat. Serverseitig sind die Tags allerdings nicht vorgeschrieben. IMAP bietet einige interessante Implementierungsdetails, so können etwa die Befehle FETCH und STORE für zahlreiche Nachrichten Manipulationen genutzt werden. In IMAP Mailboxen können Ordner definiert werden um die Nachrichten zu kategorisieren, für diese Ordner können sogar Berechtigungen vergeben werden, sodass nur bestimmte Clients auf die jeweiligen Ordner Zugriff haben. Ein wichtiges IMAP Konzept stellen die zahlreichen Flags dar. Diese können bei Nachrichten gesetzt werden beispielsweise das Answered oder Deleted flag. [KOI11]

IMAP bietet von Haus aus ein, sehr hoch entwickeltes Suchsystem, so kann z.B. nach allen Nachrichten, die beantwortet wurden und von „Robert Söllner" vor dem „15. Mai 2014" gesendet wurden, gesucht werden. Emails verbleiben bei IMAP immer auf dem Server d.h. es können beliebig, viele Clients (auch gleichzeitig) darauf zugreifen. Die einzelnen Clients können natürlich Kopien der Nachrichten anlegen um nicht ständig online sein zu müssen.

Die Änderungen werden dann bei der nächsten Verbindung synchronisiert. IMAP unterstützt durch eine Protokollerweiterung auch PUSH Emails wodurch das ständige Nachfragen des Clients ob neue Nachrichten eingelangt sind, nicht mehr notwendig ist. Nachrichten Flags können gesetzt werden, so behält man die Übersicht welche Nachrichten am Server z.B. bereits gelesen, beantwortet, gelöscht usw. wurden. IMAP bietet ebenfalls die optionale Möglichkeit nur Nachrichtenheader anzuzeigen und erst im Nachhinein - falls gewünscht - einen Download des Nachrichteninhalts einzuleiten, sowie die Möglichkeit nur Teile von Nachrichten herunterzuladen. Beispielsweise bei MIME multipart Nachrichten mit großen Multimedia Dateien und kurzem Text, nur den Text. [KOI11]

Sicherheitsüberlegungen

Bei IMAP Verbindungen findet immer eine Benutzerauthentifizierung statt. Dabei kann der IMAP eigene Mechanismus verwendet werden, bei dem allerdings das Passwort im Klartext übermittelt wird oder ein alternativer Authentifizierungsdienst wie z.B. Kerberos. Da bei IMAP alle Nachrichten zentral auf dem Server gespeichert werden und zum Client übertragen werden, sollte die Verbindung nach Möglichkeit z.B. mittels SSL oder TLS verschlüsselt werden. Dies kann der Client z.B. mit dem Kommando STARTTLS veranlassen, wodurch dann ebenfalls der in IMAP integrierte Authentifizierungsmechanismus verschlüsselt angewandt wird. Da derartige Verbindungen über Port 993 erfolgen, muss dies vom Server unterstützt werden, was z.B. vorher mit dem CAPABILITY Kommando abgefragt werden kann. Erfolgt die Verbindung verschlüsselt, so gilt IMAP als sicheres Protokoll für die E-Mail Übermittlung. Den Email Provider sollte man allerdings mit Bedacht wählen, da alle Mailboxen inklusive deren Nachrichten zentral auf dem Server des Anbieters gespeichert werden und dieser daher alle Datenschutz Bestimmungen erfüllen sollte.

3.3.1. IMAP Anonymisierungsüberlegungen

Die Anonymisierungsüberlegungen decken sich hier Großteils mit denen aus Abschnitt 3.2 Post Office Protocol Version 3 (POP3). Auch IMAP bietet mit dem STARTTLS Kommando die Möglichkeit, Verbindungen verschlüsselt aufzubauen. Ebenfalls muss IMAP über ein Anonymisierungsnetzwerk laufen, um die IP Adresse des Benutzers nicht dem Server zugänglich zu machen. Einige IMAP Server unterstützen die SASL Erweiterung: ANONYMOUS Authentification. Diese ermöglicht es Benutzern sich anonym anzumelden. Damit gewährt man jedem Benutzer Zugriff zum IMAP Server, daher wird vor der Aktivierung der ANONYMOUS Erweiterung ausdrücklich gewarnt, es sei denn der Server fungiert als öffentliche Nachrichtenplattform z.B. „Schwarzes Brett“. Im folgenden Beispiel ist die Kommunikation einer Anonymous Authentification Anmeldung zwischen Client und Server abgebildet. Der vom Client verwendete „trace“ `c2lyaGM=` kann beliebig gewählt werden und ist optional: [RFC2245]

```
S: * OK IMAP4 server ready
C: A001 CAPABILITY
S: * CAPABILITY IMAP4 IMAP4rev1 AUTH=CRAM-MD5 AUTH=ANONYMOUS
S: A001 OK done
C: A002 AUTHENTICATE ANONYMOUS
S: +
C: c2lyaGM=
S: A003 OK Welcome, trace information has been logged.
```

3.4. Simple Mail Transfer Protocol (SMTP)

Protokoll:	**Simple Mail Transfer Protocol (SMTP)**
Standards (aktuell):	RFC 2821 (2001) und RFC 5321 (2008)
Standardport(s):	TCP Port 25, TCP Port 465 (verschlüsselt) oder TCP Port 587 (Alternative für Mail-Clients)
Verbreitung:	Sehr Hoch
Schadenspotenzial:	Sehr Hoch
Funktionsgruppe:	E-Mail Protokolle
Zweck:	SMTP ist ein Protokoll, das den Versand und die Weiterleitung von E-Mail Nachrichten ermöglicht. Es wurde 1981 (RFC 788) entwickelt um auf die spezifischen Anforderungen des E-Mail Verkehrs besser eingehen zu können und diesen vom FTP-Dienst abzukoppeln. SMTP löste die damals vorherrschenden Mail Protokolle, das Mail Box Protocol sowie das Mail Transfer Protocol (MTP), die hauptsächlich auf Telnet und FTP aufbauten, ab. Die aktuelle Protokollversion stellt der RFC 2821 dar, der eine Revision des RFC 821 (1982) ist.

Funktionsweise

In der Entwicklungszeit des SMTP Protokolls musste sich das Protokoll einigen Herausforderungen stellen. Damals herrschte im Arpanet/Internet eine sehr heterogene Struktur vor, daher konnten manche Systeme miteinander nicht direkt kommunizieren. Des Weiteren gab es noch kein Domain Name System, d.h. es war unmöglich zu einer gegeben E-Mail Adresse die IP Adresse des zuständigen SMTP Servers herauszufinden. Daher setzten die Entwickler von SMTP auf Relaying. In die Nachrichten wurde einfach die Transportinformation integriert. Wenn man eine Nachricht vom Server A zum Server D senden wollte, so musste man auch die Zwischenschritte B und C inkludieren. Daraufhin wurde jeweils zwischen 2 benachbarten Knoten eine SMTP Verbindung aufgebaut und die Nachricht übermittelt und das solange, bis die Nachricht den richtigen Empfänger erreicht hatte. Diese Form der Nachrichtenübermittlung wird auch heute noch unterstützt allerdings steht seit der Einführung des Domain Name Systems (DNS) ein weitaus simplerer und effizienterer Weg zur Verfügung, Nachrichten zu übermitteln. Relaying wird heute von einigen Servern blockiert, da es auch zum Versenden von SPAM genutzt wurde/wird. DNS führte die MX (Mail-Exchanger) Records ein, was es

dem Sender einer Nachricht ermöglicht den zuständigen SMTP-Server für den Empfang einer Nachricht, für die jeweilige E-Mail Adresse ausfindig zu machen.

Daher benötigt man nur mehr zwei SMTP Verbindungs-Schritte im E-Mail Versand: einen vom Sender (Client Computer) zum SMTP-Server des Senders und einen weiteren Schritt vom SMTP-Server des Senders zum SMTP-Server des Empfängers. Diese Variante wird auch vom RFC 2821 als die präferierte Variante empfohlen. Der letzte Schritt (Der „Download" der Nachricht vom SMTP-Server des Empfängers auf den Client Rechner des Empfängers) erfolgt mittels anderer Protokolle (i.d.R. POP3 oder IMAP). Die anderen Protokolle werden benötigt, da SMTP Server permanent laufen müssen um den Empfang von Nachrichten zu ermöglichen und dies natürlich für den Client unzumutbar ist, seinen Computer permanent laufen zu lassen. Ein Schlüsselkonzept von SMTP ist es für jeden Befehl eine Rückmeldung (Replycode) zu bekommen. Abbildung 4 beschreibt den typischen Ablauf einer SMTP Verbindung: [KOS11]

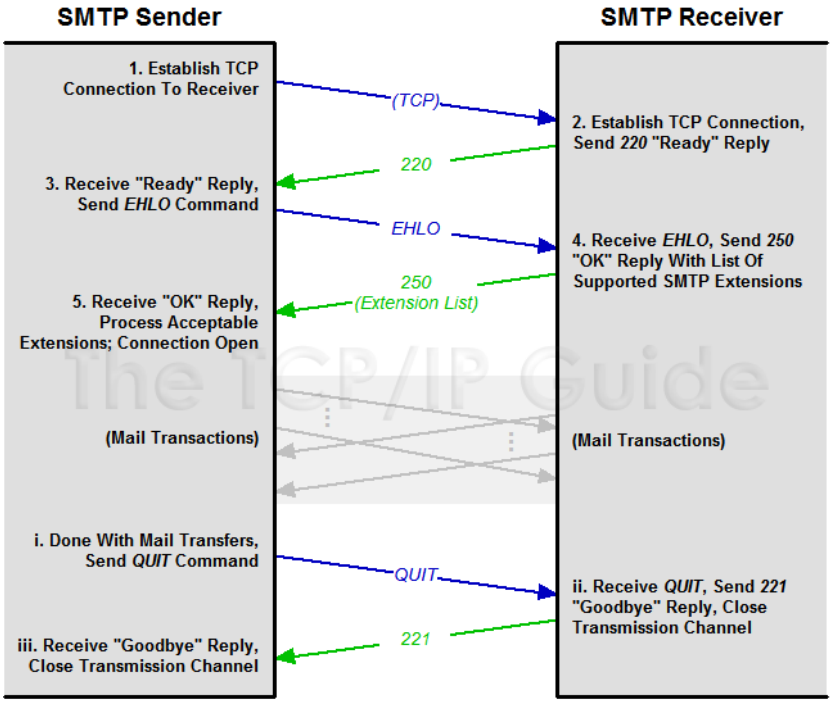

Abbildung 4: SMTP Verbindungsdiagramm

Während der Mail Transactions Phase übermittelt der Sender dem Empfänger die Absender Email-Adresse (Schritt 1), die Empfänger Email-Adresse (Schritt 2) und die eigentliche Nachricht (Header sowie Body) (Schritt 3). Dies sind drei separate Schritte. Theoretisch könnte man die Schritte 1 und 2 einsparen und nur die eigentliche Nachricht übermitteln, da die Adressen auch in der Nachricht selbst stehen. Warum benötigt man die zusätzlichen Schritte?:
SMTP verwendet das Envelope Konzept, d.h. analog zu postalischen Briefen, stehen Absender und Empfänger auf dem Briefumschlag, während der Inhalt (Content) auf den ersten Blick nicht ersichtlich ist. Einerseits ist damit eine effizientere Verteilung möglich, da der SMTP Server die „wichtigen" Informationen als erstes bekommt und damit sofort entscheiden kann

ob er die Nachricht annimmt/weiterleitet. Anderseits ist dieses Konzept ein essentieller Bestandteil der Blind Carbon Copies (BCCs). Ein Empfänger der für andere Empfänger nicht aufscheinen soll darf nicht im Nachrichtencontent stehen!

Es gibt SMTP Extensions, die verwendet werden um zusätzliche „Features“ bei SMTP zu ermöglichen, wie z.B. Pipelining (das Senden/Empfangen mehrerer Befehle gleichzeitig) oder Authentifizierungsfunktionen. Eine ausführliche Liste dieser Extension kann unter http://www.iana.org/assignments/mail-parameters eingesehen werden.

Sicherheitsüberlegungen

SMTP wurde ganz auf dem Konzept des gegenseitigen Vertrauens aufgebaut (was zur Entwicklungszeit von SMTP auch durchaus noch Sinn machte). Dies führt allerdings dazu, dass SMTP Informationen offen und unverschlüsselt in Textform übertragen werden. D.h. ein Angreifer kann die Nachrichten „mitlesen“ und kann sogar Daten ohne Probleme verändern, wenn keine weiteren Schutzmaßnahmen getroffen wurden. Eine durchaus bekannte Problematik, die sich Spammer zu Nutze machen, da SMTP Server früher prinzipiell alles weiter geleitet haben, was ihnen geschickt wurde. Heutzutage werden – abhängig von der SMTP Software – diverse Überprüfungen durchgeführt z.B. Ob die Absenderadresse korrekt oder ob die IP Adresse des Absenders in der Liste der berechtigten Absender steht, etc. Eine wirkliche Lösung der Problematik gibt es allerdings bis heute noch nicht. SMTP selbst bietet auch keine Nachrichtenverschlüsselung an, dies muss – falls gewünscht – mit Hilfe der durch MIME definierten Mechanismen erfolgen.

3.4.1. SMTP Anonymisierungsüberlegungen

In der Anfangszeit von SMTP war die Mehrheit der SMTP Server als „Open Relay“ konfiguriert. Es wurden weder IP Adressen noch Benutzer überprüft und alle Nachrichten zur Weiterversendung einfach angenommen. Diese „anonyme“ Vorgehensweise führte bald dazu, dass Spammer die SMTP Server für ihre unerwünschten Zwecke ausnutzen. Dies wiederum hatte zur Folge, dass die Betreiber der SMTP Server von dieser Praxis Abstand nahmen.

Heutzutage nehmen SMTP Sever Nachrichten nicht mehr von beliebigen Absendern entgegen, sondern sie akzeptieren nur noch diejenigen Nachrichten, bei denen sichergestellt ist, dass sie zum jeweiligen SMTP Server passen. Dies geschieht z.B. anhand von IP/Domain Kennungen und Benutzeranmeldedaten. Diese Praxis erschwert die anonyme Verwendung von SMTP erheblich. Daher wurden Anonymous Remailer entwickelt, die einerseits technisch die Anonymisierung ermöglichen und andererseits auch noch darüber hinausgehende Möglichkeiten wie beispielsweise die inhaltliche Filterung von E-Mails ermöglichen. Um die Anonymität weiter zu steigern verkettet man optimaler Weise mehrere Anonymous Remailer und verwendet verschlüsselte Verbindungen zwischen den Remailern. Falls die Anonymous

Remailer selbst keine Weiterleitungsdaten speichern und der Absender den Inhalt der Nachricht möglichst anonym gestaltet hat, so ist es nahezu unmöglich für den Empfänger der Nachricht den wahren Absender zu ermitteln.

4. Cypherpunk Anonymous Remailer (Type I)

4.1. Implementierung

Die Software wurde in der Programmiersprache Java SE Version 1.7 verfasst.

Es werden drei Hauptframeworks verwendet mit deren Hilfe die Remailer Funktionalitäten realisiert werden. Dies sind folgende Frameworks:

- **Stanford Named Entity Recognizer (NER) für die Namensfindung**
  (http://www-nlp.stanford.edu/software/CRF-NER.shtml)
  Lizenz: GNU General Public License (v2 or later)
- **SubEtha SMTP für den E-Mail Empfang mittels SMTP**
  (https://code.google.com/p/subethasmtp/)
  Lizenz: Apache License, Version 2.0
- **Simple Java Mail für das Senden von E-Mail Nachrichten**
  (http://code.google.com/p/simple-java-mail/)
  Lizenz: Apache License, Version 2.0

4.1.1. Pakete- und Klassenüberblick

Neben den zahlreichen Klassen der Frameworks werden natürlich auch weitere Klassen benötigt, die diese Frameworks miteinander verbinden und die Remailer spezifischen Eigenschaften implementieren. Diese selbst erstellten Codeteile für den Anonymous Remailer gliedern sich in folgende Pakete und Klassen:

- **main**
  - MainGUI.java – Enthält Main Methode und GUI der Anwendung
  - StarterClass.java – Enthält Methoden zum initialisieren und stoppen der Anwendung bzw. der Anwendungskomponenten
- **inputmgmt**
  - BasicSMTPServer.java – Klasse implementiert den SMTP Server und steuert diesen
  - HeaderVectorObject.java – Speicherformat für Email Dummy Header
  - MyMessageHandlerFactory.java – Enthält Logik für die Bearbeitung von empfangenen Email Nachrichten
- **nlpprocessing**
  - NLPAnonymizer.java – Hauptfunktionen des Stanford Natural Language Processors, wird verwendet um Text zu anonymisieren.

- **outputmgmt**
  - MessageBuffer.java – Verwaltet eine Kette von Email Nachrichten und verzögert deren Exekution bis zum gewünschten Zeitpunkt
  - MessageObject.java – Speicherformat für Email Nachrichten
  - OutputHandler.java – Zuständig für das Senden der Email Nachrichten

Die folgende Abbildung veranschaulicht die Zusammenhänge auf Package Level (ohne das NLP-Framework und das SimpleJavaMail-Framework):

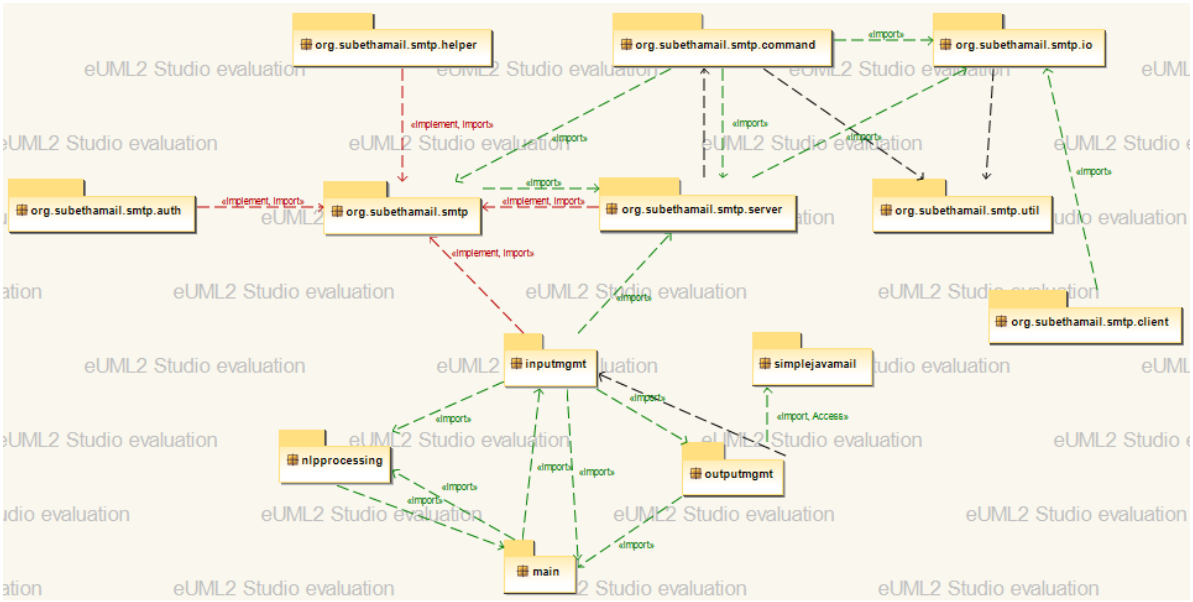

Abbildung 5: Package Dependency Graph

Das gleiche Diagramm wäre als Abhängigkeitsgraph auf Klassenlevel aufgrund der Komplexität nicht mehr darstellbar, daher wird auf diese Darstellungsform verzichtet. Abbildung 6 soll die Zusammenhänge (ohne die Frameworks) auf Klassen Level verdeutlichen. Die Pfeile zwischen den Klassen sollen den zeitlichen Ablauf verdeutlichen:
Email Nachrichten werden vom Handler in der MyMessageHandlerFactory empfangen. Dann mit Hilfe der NLPAnonymizer Klasse anonymisiert. Anschließend wird ein MessageObject angelegt und dieses im MessageBuffer verwaltet, bis es zum Zeitpunkt der Versendung mit Hilfe der OutputHandler Klasse versandt wird.

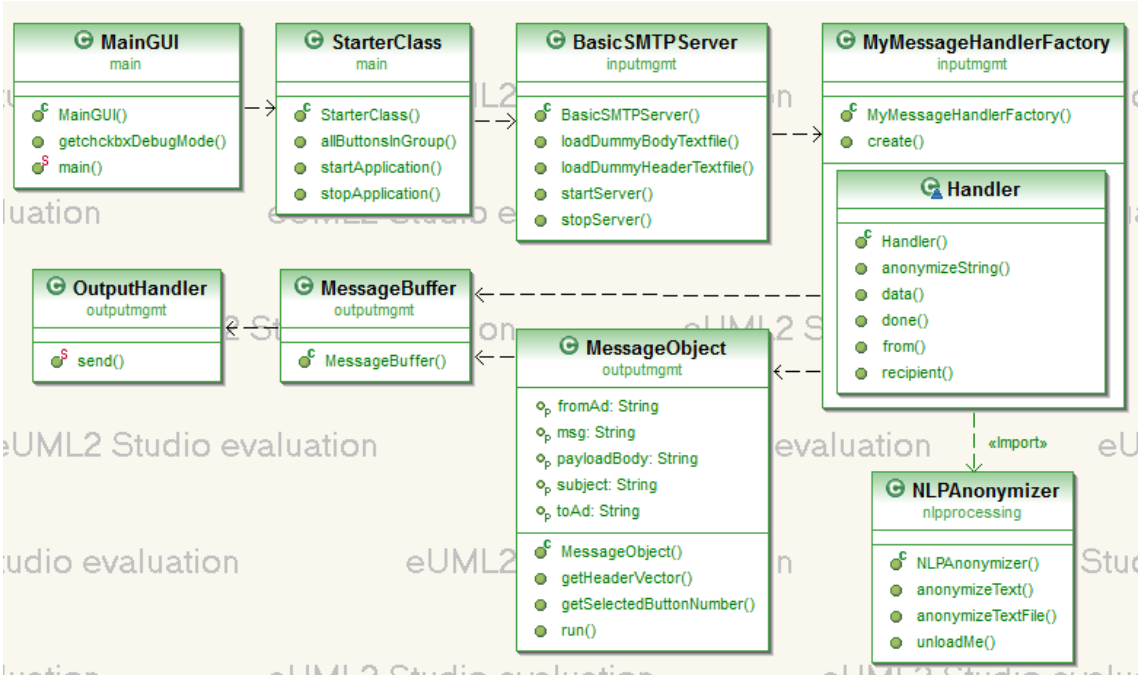

Abbildung 6: Class Activity Flow

4.2. Natural-Language-Processing (NLP)

Unter Natural-Language-Processing versteht man die Wissenschaft, die sich mit der Erkennung und Verarbeitung von menschlicher Sprache in technischen Systemen beschäftigt. Die Forschung kombiniert dabei Erkenntnisse aus der Informatik, der Linguistik und der Forschung an Künstlicher Intelligenz. Die Anfänge der modernen Forschung sind auf die 1950er Jahre zurück zu führen, wo unter anderem Alan Turing den Artikel „Computing Machinery and Intelligence“ veröffentlichte. Dieser Artikel ist heutzutage besser bekannt als „Turing Test“. Bei diesem Test soll ein menschlicher Fragesteller über ein Terminal rein textuell (d.h. ohne Sicht- oder Sprechverbindung) Fragen an zwei Probanden stellen. Einer der Probanden ist eine Künstliche Intelligenz („Computer“) der andere ein Mensch. Ist es dem Fragesteller nach dem Ablauf einer gewissen Zeit (in der Annahme sind es fünf Minuten) noch immer nicht möglich zu unterscheiden wer von den Probanden der Mensch und wer die künstliche Intelligenz ist, so gilt der Turing Test als bestanden. Der Test zielt vor allem auf die künstliche Intelligenz eines Systems ab, jedoch ist Natural-Language-Processing ein essentieller Bestandteil des Tests, da die textuellen Eingaben vom System erfasst und „verstanden“ werden müssen um „passende“ Antworten geben zu können.

Maschinelles Lernen („Machine learning“)

Die meisten modernen NLP Algorithmen arbeiten nach der Methode des maschinellen Lernens („Machine learning“). Frühere NLP Algorithmen beruhten zumeist auf der Eingabe einer großen Zahl von Regeln um die natürliche Sprache regelbasiert abbilden zu können. Aufgrund der Komplexität und Wandelbarkeit unserer Sprache und des erheblichen Aufwands

kam rasch die Idee auf, die Computer diese Regeln selbst lernen zu lassen. Dies wird mit der Analyse von großen Mengen an natürlich Texten, die jedoch zuvor schon manuell vorklassifiziert wurden, erreicht („Corpora“). Durch die Ermittlung der statistischen Abweichung, lassen sich so automatisiert gewichtete Regelsätze erstellen, die anschließend für die Erkennung ähnlicher Muster verwendet werden können.

Named-Entity-Recognition

Natural-Language-Processing besteht aus einer Vielzahl an Unterthemen, die sich mit der Forschung an Teilgebieten des Natural-Language-Processings beschäftigen. Das für diese Software relevante Teilgebiet nennt sich „Named-Entity-Recognition“ und ist wiederum ein Teilgebiet der „Information Extraction“. Die „Named-Entity-Recognition“ abgekürzt NER, beschäftigt sich mit der Erkennung von Namen/Entitäten in natürlich sprachlichen Texten. Anschließend müssen die Entitäten einer semantischen Kategorie zugewiesen werden. Diese Zuweisung ist stark vom jeweiligen Kontext, in der sich die Entität befindet, abhängig. Zum Beispiel: Wird in einem Text von einem „Mr. Bernstein“ gesprochen handelt es sich um eine Person, wird jedoch von einem „Bernstein college“ gesprochen, so handelt es sich um eine Organisation. Ein weiteres Beispiel: Liest man in einem Text von „Berlin Medical Center“ so handelt es sich um eine Organisation. „The Medical Center in Berlin“ handelt jedoch von einer (nicht näher bestimmten) Organisation „The Medical Center“ und einem Ort „Berlin“. Diese kontextbasierte Kategorisierung ist hoch komplex und erfordert sehr spezifische Software.

Im Bereich der Information Extraction gibt es zahlreiche Software Programme, die fast alle eines der folgenden drei statistischen „hidden state“ Sequenzmodelle oder Varianten davon einsetzen:

- Hidden Markov Models (**HMM**s) (beschrieben von: Leek, 1997; Freitag und McCallum, 1999) [RGM05]
- Conditional Markov Models (**CMM**s) (beschrieben von: Borthwick, 1999) [RGM05]
- Conditional Random Fields (**CRF**s) (beschrieben von: Lafferty et al., 2001) [RGM05]

Es war Teil dieser Masterarbeit ein geeignet NER Framework zu finden, welches sehr gute Erkennungsraten aufweist, effizient arbeitet und dabei die Software nicht unnötig aufbläht. Nach einigen Vortests habe ich mich für den Stanford Named Entity Recognizer (NER) entschieden:
http://www-nlp.stanford.edu/software/CRF-NER.shtml

4.2.1. Stanford Named Entity Recognizer (NER)

Stanford NER ist ein so genannter CRFClassifier. Die Software implementiert ein linear verkettetes „Conditional Random Field (CRF)"-Sequenzmodell. Conditional Random Fields (CRFs) sind eine Klasse von statistischen Modellierungsmethoden, die vor allem im Bereich der Mustererkennung und des Machine Learnings eingesetzt werden. Sie werden dort zur Segmentierung von Sequenzen verwendet. Stanford NER verwendet es, um die Eingabedaten zu klassifizieren und erzeugt für jede Eingabesequenz eine gleich lange Ausgabesequenz. Die zugehörige Funktion lautet P(y|x), das bedeutet in Abhängigkeit von der Beobachtungssequenz x wird die bedingte Wahrscheinlichkeit für die Labelsequenz y bestimmt. Ein CRF kann in jedem Zustand auf die gesamte Information der Beobachtungssequenz zugreifen, da CRFs ungerichtete Modelle sind. Diese Eigenschaft unterscheidet CRFs von gerichteten Modellen – beispielsweise den Hidden-Markov-Modellen (HMM) - die jeweils nur auf die aktuelle Beobachtung zugreifen können.

CRFs zählen zur Klasse der „discriminative models" und haben im Gegensatz zu generativen Modellen den Nachteil, dass man mit ihnen keine Stichproben aus der gemeinsamen Verteilung von x und y generieren (berechnen) kann. Aufgabenstellungen wie beispielsweise das Klassifizieren von Texten benötigen diese Eigenschaft jedoch nicht. Da die gemeinsame Verteilung nicht berechnet werden muss, sind CRFs sehr perfomant. Aufgrund der erwähnten Einschränkung ist es jedoch notwendig die Stichproben mittels eines Sampling Algorithmus zu finden. Es muss dabei nur die bedingte Wahrscheinlichkeit für jede Position in der Sequenz berechnet werden – nicht der konkrete Wert. Klassische CRF-Modelle betrachten nur ein sehr schmales Fenster an Variablen (local-strucure/evidence). Dafür werden beispielsweise der Viterbi- oder der Beamsearch-Algorithmus verwendet um bei einer gegebenen Beobachtungssequenz x die optimale Labelsequenz y zu finden. Diese Algorithmen sind im Stanford NER (zu Vergleichszwecken) implementiert. [RGM05] Es folgt eine kurze Erklärung der beiden Algorithmen:

Viterbi-Algorithmus

Der Viterbi-Algorithmus wird im Bereich der Künstlichen Intelligenz für die Erkennung von versteckten Zuständen eingesetzt. Er ist für die Erkennung von Mustern ausgelegt und sein Hauptvorteil ist die „nur" linear steigende Komplexität bei Vergrößerung der Eingabe-Sequenzlänge (bei anderen Algorithmen steigt die Komplexität exponentiell mit der Eingabe-Sequenzlänge).

Beamsearch-Algorithmus

Der Beamsearch-Algorithmus generiert einen Suchbaum. Bei jeder Stufe des Baums werden jeweils die möglichen Nachfolge-Zustände als nächste Stufe angehängt. Im Unterschied zur normalen Breitensuche werden die Nachfolge-Zustände nach den heuristischen Wahrschein-

lichkeiten geordnet und nur diejenigen Zustände weiterverfolgt, die die höchste Wahrscheinlichkeit haben. Die Anzahl, wieviel Zustände pro Stufe weiterverfolgt werden sollen lässt sich angeben und wird „beam width“ genannt. Hauptvorteil ist, dass sich der Speicherplatzverbrauch selbst bestimmen lässt. Nachteil ist, dass jeweils die erste Lösung gefunden wird und nicht notwendigerweise die Beste und das speziell bei geringer „beam width“ eventuell gar keine Lösung gefunden wird („non-conpleteness“).

„Gibbs“-Sampling

Im Stanford NER ist zusätzlich der Gibbs-Sampling Algorithmus implementiert. Er ermöglicht es das Betrachtungsfenster auf weitere Variablen auszudehnen (non-local-structure/evidence). Das folgende Beispiel soll den Mehrwert verdeutlichen, den eine Einbeziehung der „non-local-structure/evidence“ und damit Gibbs-Sampling im Stanford NER bietet. Man betrachte beispielsweise folgenden Satzauszug „… the news agency *Tanjug* reported … airport, *Tanjug* said.“ [RGM05]. Beim zweiten Vorkommen des Wortes „Tanjug“ ist es nicht ersichtlich ob es sich um eine Person oder eine Organisation handelt. Lässt man jedoch die Information aus einem bereits vorherigem Vorkommen des Wortes „Tanjug“ mit einfließen so erhält man genügend Information um „Tanjug“ als Organisation einstufen zu können („the news agency *Tanjug* reported …“). Abbildung 7 veranschaulicht das beschriebene „Label Consistency Problem“ [RGM05]:

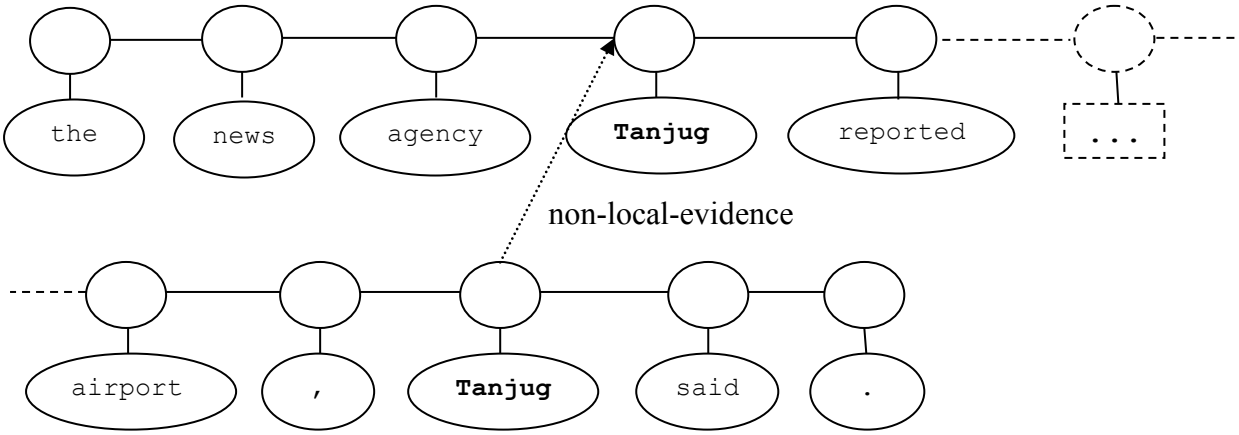

Abbildung 7: Label Consistency Example

Grundlagen des „Gibbs“-Sampling

Der Gibbs-Sampling-Algorithmus ist ein Sonderfall des Metropolis-Hastings-Algorithmus, welches eine praktische Anwendung des Markov-Chain-Monte-Carlo-Verfahrens (kurz MCMC-Verfahren) darstellt. Mit Hilfe von Gibbs-Sampling kann die gesuchte gemeinsame Verteilung der Zufallsvektoren approximiert werden um den wahrscheinlichsten Statusübergang vorherzusagen. Mathematisch kurz zusammengefasst:

Gegeben $\theta_1, \theta_2 \sim f(\theta_1, \theta_2)$ so lassen sich sampeln $f(\theta_1 | \theta_2)$ und $f(\theta_2 | \theta_1)$.

Beginnend mit dem Initialwert $(\theta_1^{(0)}, \theta_2^{(0)})$ lautet der Gibbs-Sampling Algorithmus:

1. sample $\theta_1^{(j)} \sim f(\theta_1 | \theta_2^{(j-1)})$ und dann
2. sample $\theta_2^{(j)} \sim f(\theta_2 | \theta_1^{(j)})$

Es wird eine Folge von Stichproben aus einer gemeinsamen Wahrscheinlichkeitsverteilungen mehrerer Zufallsvariablen ermittelt. Das Ziel ist es, die unbekannte gemeinsame Verteilung zu approximieren. Man wählt eine Variable aus und berechnet den Wahrscheinlichkeitswert in Abhängigkeit von den Werten aller anderen Variablen. Die Werte der anderen Variablen bleiben in diesem Iterationsschritt unverändert. Dies wiederholt man so lange bis für jede Variable ein Wahrscheinlichkeitswert in Abhängigkeit von allen anderen Variablen vorliegt. Aus der entstehenden Folge von Stichprobenvektoren lässt sich eine Markov-Kette herleiten. Diese kann als gewichteter Graph visualisiert werden, wo jeder Zustandsübergang mit der jeweiligen Wahrscheinlichkeit gewichtet ist.

Stellt man sich diesen Graphen nun als Ansammlung von Token (Wörter) und Klassifizierungsmöglichkeiten vor, so lässt sich mit Hilfe der errechneten Wahrscheinlichkeiten das wahrscheinlichste Klassifizierungsergebnis für das jeweilige Token (Wort) bestimmen. Ist dieses bestimmt, so wird das entsprechende Klassifikationsergebnis als inline-XML mit dem Token verknüpft. Die zusammen gefügten Tokens ergeben dann den klassifizierten Text. Dieser wird als Endergebnis zurückgegeben.

Classifier-Varianten

Die im Anonymous Remailer für das Natural-Language-Processing verwendeten Classifier versuchen mit Hilfe errechneter Wahrscheinlichkeiten, den Eingabetext, in jeweils passende Kategorien (z.B. Namen, Orte, Organisationen, etc.) einzuordnen. Die eingehenden E-Mail Nachrichten werden vereinfacht beschrieben durch den verwendeten Classifier in Tokens zerteilt und diese werden dann wiederum im Classifier mit den vorberechneten Werten verglichen. Diese vorberechneten Werte werden in aufwendigen Prozessen anhand von Textsamples berechnet. Je mehr Textsamples ein Classifier „bekommt“, desto genauer werden die vorberechneten Wahrscheinlichkeiten. Es werden Wortstellung, Grammatik, Satzfluss und viele weitere Faktoren in die Berechnung mit einbezogen. Die Verteilung der Faktoren ist von der verwendeten Sprache abhängig. Daher werden die Classifier immer für eine Zielsprache „trainiert“. Der Anonymous Remailer enthält vorinstalliert die Classifier für die Sprachen Englisch und Deutsch.

In der Anonymous Remailer Software stehen zwei Sprachen für die bereits eingebundenen Classifier zur Auswahl: Der englische Classifier ist ein 3-class-Classifier und klassifiziert nach „Person“, „Organisation“ und „Ort“. Die deutschen Classifier sind 4-class-Classifier und klassifizieren nach „Person“, „Organisation“, „Ort“ und „Verschiedenes“. „Verschiedenes“ ist eine Mischkategorie in die diejenigen Ergebnisse aufgenommen werden die für mehr als eine Kategorie passend sind und daher nicht eindeutig zugeordnet werden können. Der englische Classifier wurde mit Text aus Webseiten trainiert. Der deutschsprachige ebenfalls, daneben gibt es auch noch einen deutschsprachigen Classifier, der mit Zeitungstexten trainiert wurde (dieser ist im untenstehenden Codelisting auskommentiert). Es können jederzeit weitere Classifier (Sprachen) mit geringfügigen Codeanpassungen hinzugefügt werden. Voraussetzung ist, dass diese für das Stanford NER entwickelt wurden. Ein Algorithmus zum Wechseln des verwendeten Classifiers, während der Anonymous Remailer gestartet ist, ist im Anonymous Remailer enthalten. Aufgrund des hohen Arbeitsspeicher-Platzverbrauchs der Classifier können leider auf gewöhnlichen Systemen nicht mehrere Classifier gleichzeitig verwendet werden.

```
public class NLPAnonymizer {

	private String serializedClassifierEnglish =
	"/classifiers/english.all.3class.distsim.crf.ser.gz";

	// Warning at least 800MB heap size needed,
	private String serializedClassifierGerman =
	"/classifiers/dewac_175m_600.crf.ser.gz";
	// "/classifiers/hgc_175m_600.crf.ser.gz" for newspaper trained classifier
...
}
```

Im Kapitel 4.4.1 Testfälle und Erkennungsergebnisse wird ebenfalls darauf eingegangen wie die Classifier mit Texten umgehen, die in einer Sprache verfasst sind, die ungleich derjenigen Zielsprache sind für die der jeweilige Classifier trainiert wurde.

4.2.2. Classifier-Training mit eigenen Daten

Es ist möglich den Stanford Classifier mit eigenen Daten zu trainieren. Dazu werden die Klassen *CRFClassifier* und *NERFeatureFactory* benötigt. Die Verwendung der Klassen ist in den jeweiligen Javadoc Dateien dokumentiert. Die Trainingsdaten müssen aus zwei mittels Tab getrennten Spalten bestehen. In der linken Spalte soll „word“ genannt werden und in ihr stehen die Eingabetokens (Wörter). Die rechte Spalte soll „answer“ genannt werden und in ihr muss die jeweilige Klassifikation (NER-Klasse) für jedes Wort der linken Spalte stehen. Diese Benennung der Spalten, ebenso wie die Angabe welches Modell und welche Features angewandt werden sollen erfolgt in einem Property file. Es folgt ein Beispiel für eine einfache NER-properties Datei (test.prop):

```
    trainFile = training-data.tok
```

```
serializeTo = ner-model.ser.gz
map = word=0,answer=1

useClassFeature=true
useWord=true
useNGrams=true
noMidNGrams=true
maxNGramLeng=6
usePrev=true
useNext=true
useSequences=true
usePrevSequences=true
maxLeft=1
useTypeSeqs=true
useTypeSeqs2=true
useTypeySequences=true
wordShape=chris2useLC
useDisjunctive=true
```

Die Klasse nlp.process.PTBTokenizer kann verwendet werden um eine normalen Text Datei mit Trainingsdaten in das für das Classifier-Training richtige Format zu bringen. Es wird eine so genannte TOK-Datei generiert. In der Datei steht dann jedes Wort in einer eigenen Zeile. In der „answer“-Zeile muss dann noch manuell die jeweilige Klassifikation eingetragen werden. Ein Beispiel shell-Kommando für diesen Vorgang lautet:

```
java -cp stanford-ner.jar edu.stanford.nlp.process.PTBTokenizer
exampleinputfile.txt > exampletokenfile.tok
```

Das Training des Classifiers wird beispielsweise mit folgendem shell-Kommando eingeleitet:

```
java -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -prop
test.prop
```

Dadurch wird - sobald der Vorgang abgeschlossen ist - ein passendes NER-Modell an dem in der Property Datei (test.prop) angegeben Speicherort (Dateiname im Beispiel: ner-model.ser.gz) generiert. Um zu überprüfen wie erfolgreich das Training war, kann nun ein Vergleichstest mit dem Classifier gestartet werden. Es folgt ein Beispiel für das zugehörige shell-Kommando:

```
java -cp stanford-ner.jar edu.stanford.nlp.ie.crf.CRFClassifier -loadClassifier
ner-model.ser.gz -testFile exampletokenfile.tok
```

Die erste Spalte der erzeugten Datei zeigt das Eingabewort. Die zweite Spalte zeigt die manuell-eingegebene, korrekte Klassifikation („gold-answer“) und die dritte Spalte zeigt die vom Classifier ermittelte Klassifikation („answer“). Auf diese Weise können gezielt die Schwachpunkte in der Erkennungsleistung des Classifiers analysiert werden und diese können durch angepasste Trainingsdaten reduziert/behoben werden. Am Ende des Testfiles ist jeweils die Gesamt-Präzision in Prozent pro Klassifikations-Kategorie angegeben. Eine ausführliche Anleitung inklusive herunterladbarer Codebeispiele findet man auf der Stanford-NER Webseite im Bereich „CRF-FAQ“: http://nlp.stanford.edu/software/crf-faq.shtml#a

4.2.3. Anonymous Remailer „Regular-Expressions“

Die Liste der Classifier kann im Anonymous Remailer beliebig ergänzt werden. Wenn sich jedoch die Erkennungs-Tags verändern, so muss auch die dahinter liegende Logik zur Auswertung geändert werden. Im folgenden Codelisting sind diejenigen Regular Expressions gelistet, die den bereits klassifizierten Text nach den zuvor spezifizierten Tags scannen. Es werden dabei alle erkannten Tags jeweils einer Kategorie zugeordnet und mit vordefinierten Texten ersetzt um den Text zu anonymisieren:

```
// regex code (?s) enables multi line mode, otherwise multi line recognition would
fail
String regexPatternP = "(?s)(<PERSON>)(.*)(</PERSON>)";
String regexPatternO = "(?s)(<ORGANIZATION>)(.*)(</ORGANIZATION>)";
String regexPatternL = "(?s)(<LOCATION>)(.*)(</LOCATION>)";
String regexPatternPG = "(?s)(<I-PER>)(.*)(</I-PER>)";
String regexPatternOG = "(?s)(<I-ORG>)(.*)(</I-ORG>)";
String regexPatternMISCG = "(?s)(<I-MISC>)(.*)(</I-MISC>)";
String regexPatternLG = "(?s)(<I-LOC>)(.*)(</I-LOC>)";

if (usedClassifier == 0) { // English classifier used
      if (removePersons == true)
classifiedString = classifiedString.replaceAll(regexPatternP, personReplacement);
      if (removeOrgs == true)
classifiedString = classifiedString.replaceAll(regexPatternO, orgReplacement);
      if (removeLocs == true)
classifiedString = classifiedString.replaceAll(regexPatternL, locationReplacement);

} else { // German classifier used
      if (removePersons == true)
classifiedString = classifiedString.replaceAll(regexPatternPG, personReplacement);
      if (removeOrgs == true)
classifiedString = classifiedString.replaceAll(regexPatternOG, orgReplacement);
      if (removeLocs == true)
classifiedString = classifiedString.replaceAll(regexPatternLG,
locationReplacement);
      if (removeMisc == true)
classifiedString = classifiedString.replaceAll(regexPatternMISCG, miscReplacement);
}
```

Die Klassifizierung wird mit folgenden Methodenaufruf in der Klasse **nlpprocessing.NLPAnonymizer** gestartet:

```
String classifiedString = classifier.classifyWithInlineXML(text);
```

„classifier“ ist ein Objekt der Klasse **edu.stanford.nlp.ie.crf.CRFClassifier**. Diese erbt Methoden von der abstrakten Klasse edu.stanford.nlp.ie.AbstractSequenceClassifier<IN extends CoreMap>

In der Klasse ist die Methode „classifyWithInlineXML“ spezifiziert:

```
public String classifyWithInlineXML(String sentences) {
    return classifyToString(sentences, "inlineXML", true);
}
```

Diese stellt aber nur eine vorgeschaltete Methode für die eigentliche Klassifizierungsmethode „classifyToString“ dar und übergibt die Art der gewünschten Kennzeichnung, da mehre mögliche Arten unterstützt werden. Die Methode erwartet als Eingabeformat einen plaintext oder

einen xml String. Um den Eingabestrom zu lesen wird die Hilfsklasse edu.stanford.nlp.sequences.DocumentReaderAndWriter<CoreLabel> verwendet. Der Classifier splittet den Eingabetext in Tokens auf und behandelt dabei jeden Satz wie ein eigenes Dokument. Die Satzgrenzen werden heuristisch bestimmt. Als Ausgabeformat können folgende Arten ausgeben werden:

Beispieltext: „Bill Smith died“

- Schrägstriche ‚/‘ (z.B. Bill/PERSON Smith/PERSON died/O ./O)
- Eingebettetes XML (z.B. <PERSON>Bill Smith</PERSON> died)
- Eigenständige XML Tags (z.B. <wi num="0" entity="PERSON">Bill</wi> <wi num="1" entity="PERSON">Smith</wi> <wi num="2" entity="O">died</wi>)

Nur bei eingebettetem XML werden die Tokens genau an der Stelle in den Ausgabetext eingefügt, wie sie auch im Eingabetext vorkommen. Die anderen beiden Formate zeigen die Tokens in der Reihenfolge wie sie durch den Normalisierungsprozess im Tokenizer entstanden ist. Für den Anonymous Remailer wird eingebettetes XML verwendet, da es einfach zu filtern ist und den Textfluss nicht stört. Des Weiteren kann als Eingabeparameter noch eingegeben werden ob Abstände entfernt oder beibehalten werden sollen. Bzw. ob jeder XML Tag in einer separaten Zeile stehen soll. Für den Anonymous Remailer ist es zweckmäßig die original Abstände beizubehalten und die XML Tags in den normalen Textfluss zu integrieren.

```
public String classifyToString(String sentences, String outputFormat, boolean
preserveSpacing)
```

Innerhalb der classifyToString Methode kommt folgendes Codesegment vor:

```
edu.stanford.nlp.objectbank.ObjectBank<List<IN>> documents =
makeObjectBankFromString(sentences, plainTextReaderAndWriter);
```

Der Textstring wird mit dem Tokenizer in Tokens („Wörter“) aufgeteilt und anhand heuristisch bestimmter Kontextgrenzen in Documents („Dokumente“) zusammengefasst. Die Dokumentliste wird nun durch iteriert. Jedes Dokument wird wiederum an die classify()-Methode übergeben.

```
for (List<IN> doc : documents) {
      List<IN> docOutput = classify(doc);
...
}
```

Klassifikations-Algorithmen

Die classify()-Methode führt je nach voreingestelltem Flag entweder eine Klassifizierung mit dem Gibbs-, dem Viterbi- oder dem Beamsearch-Sampling-Algorithmus durch, für eine nähere Beschreibung der Algorithmen siehe 4.2.1 Stanford Named Entity Recognizer (NER).

```
Klasse: edu.stanford.nlp.ie.crf.CRFClassifier
  @Override
  public List<IN> classify(List<IN> document) {
    if (flags.doGibbs) {
      try {
        return classifyGibbs(document);
      } catch (Exception e) {
        System.err.println("Error running testGibbs inference!");
        e.printStackTrace();
        return null;
      }
    } else if (flags.crfType.equalsIgnoreCase("maxent")) {
      return classifyMaxEnt(document);
    } else {
      throw new RuntimeException("Unsupported inference type: " + flags.crfType);
    }
  }
```

Anwendung des „Gibbs"-Sampling

Per Default wird bei den im Anonymous Remailer verwendeten Classifiern der Gibbs-Algorithmus verwendet, da dieser aufgrund der besten Klassifikationsleistung (laut [RGM05] im Mittel 1,3% akkuratere Klassifikationsleistung im Vergleich zu Viterbi) die von den Entwicklern des „Stanford-NER-Frameworks" getroffene Voreinstellung ist:

Eingangsmethode:

```
public List<IN> classifyGibbs(List<IN> document)
```

Rekursiver Aufruf bis die Lösung gefunden wurde:

```
public List<IN> classifyGibbs(List<IN> document, Triple<int[][][], int[],
double[][][]> documentDataAndLabels)
```

Angewandt auf den Anonymous Remailer ergibt sich folgende Vorgangsweise: Der Eingabetext wird zuerst von einem Tokenizer in einzelne Tokens (z.B. Wörter) aufgeteilt. Anschließend wird mit einem POS-Tagger (Part-of-speech) jedes Token syntaktisch generalisiert und anhand sprachwissenschaftlicher Theorien in Wortklassen (engl. Part-of-speech) zugeordnet. Diese beschreiben die grammatikalische Funktion des betrachteten Wortes. Für die Wortklassen wird mit Hilfe der bereits beschriebenen Algorithmen das zugehörige Klassifikationsergebnis ermittelt. An dieser Stelle wird das CRF-Modell benötigt. Im Anonymous Remailer gibt es die Klasse edu.stanford.nlp.ie.crf.CRFClassifier, diese ist der Einstiegspunkt in das NLP Framework, für den Klassifikations-Vorgang. Dabei kommen sowohl Tokenizer, POS-Tagger als auch das CRF-Modell und Gibbs-Sampling zum Einsatz.

4.3. Konfiguration und Administration

Der Anonymous Remailer wurde in Java geschrieben und ist daher auf jedem System einsatzfähig, welches Java unterstützt. Die Software liegt wahlweise als Sourcecode oder als executable java file vor. Im letzteren Fall lässt sie sich einfach mittels Doppelklick auf die Datei anonymous_remailer_soellner.jar starten:

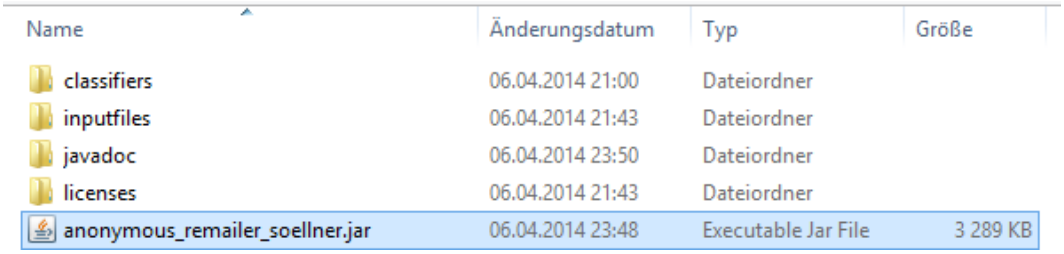

Abbildung 8: Software Folder Structure

Im „Main“ Tab ist der aktuelle Server Status ersichtlich. Mit START kann der Server gestartet werden und mit der gleichen Schaltfläche wieder gestoppt werden. Sobald der Server läuft werden am „Display“ Kontrollinformationen des Servers dargestellt. Standardmäßig werden nur wenige Informationen angezeigt um das „Display“ übersichtlich zu halten. Wenn gewünscht, dann können mittels der Checkbox „detailed output“ auch Informationen auf Nachrichtenebene mit angezeigt werden. Das „Display“ bietet einen Bildlauf, sollte es unüberischtlich werden, so kann der aktuelle Inhalt mittels der Schaltfläche „Clear“ gelöscht werden. Die folgende Abbildung zeigt das „Main“ Tab:

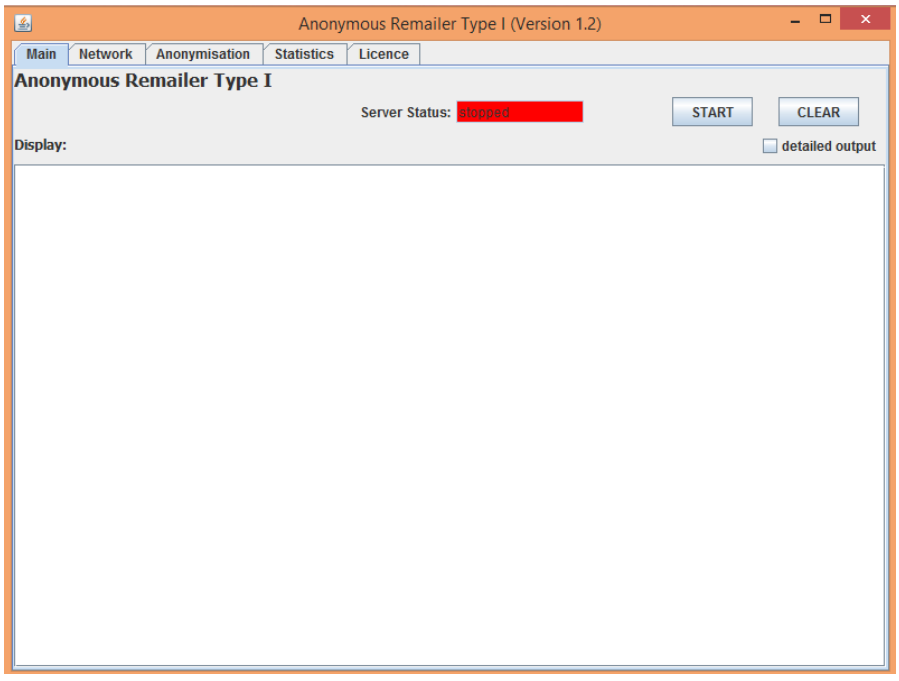

Abbildung 9: Anonymous Remailer „Main“ tab

Im „Network“ Tab kann der TCP Port konfiguriert werden, auf dem der Server auf eingehende SMTP Nachrichten wartet. Im Bereich „Outgoing Connections“ kann die „SMTP Server Url“ und ggf. „Username“ und „Password“ des SMTP Servers eingegeben werden, der die E-Mail Versendung tatsächlich durchführt. Desweiteren muss noch der TCP Port des Servers angegeben werden und die verwendete Verschlüsselung (es wird zwischen keiner Verschlüsslung/aktiviertem SSL und aktiviertem TLS unterschieden). Es wird hier bewusst ein eigener Server (vorzugsweise bei einem öffentlichen Provider) verwendet um die SPAM Problematik bei Betrieb eines eigenständigen Servers zu umgehen. Im Bereich „Server information“ sollen der gewünschte Name und eine beliebige E-Mail Adresse eingegeben werden, die als FROM Adresse der vom Remailer gesendeten E-Mail Nachrichten aufscheint. Die folgende Abbildung zeigt das „Network“ Tab:

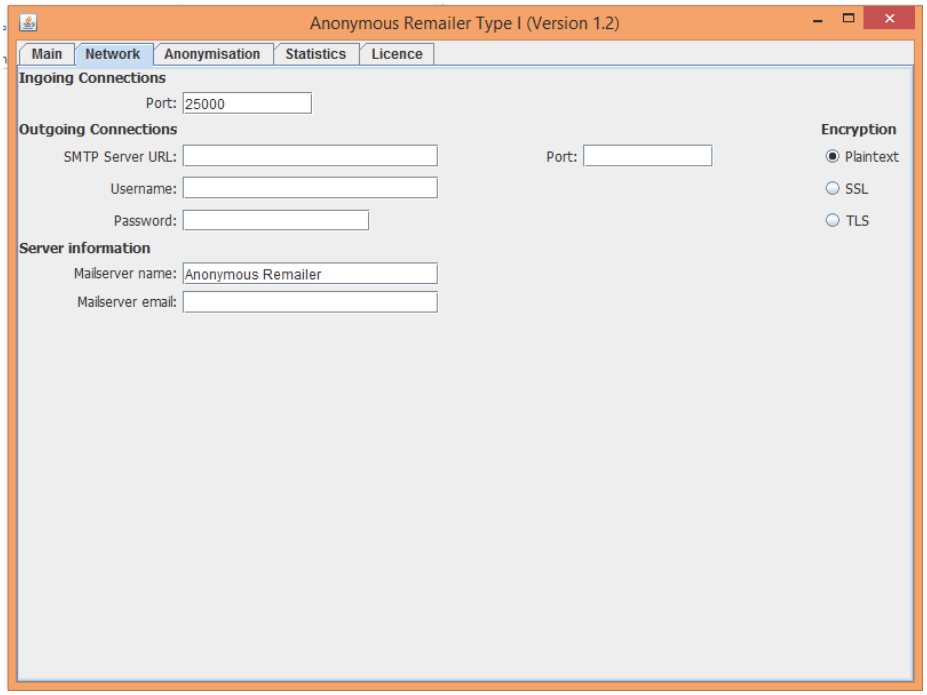

Abbildung 10: Anonymous Remailer „Network“ tab

4.3.1. Anonymisation Tab

Im „Anonymisation“ Tab werden alle Parameter konfiguriert, die für die Anonymisierung relevant sind. Es kann ausgewählt werden ob Personen (Namen), Organisationen, Orte oder keiner Kategorie zuordenbare verdächtige Informationen (diese Funktionalität bietet nur der deutschsprachige Classifier) gefunden werden sollen. Bei „Replacement text“ kann jeweils der Text eingegeben werden, mit dem das gefundene Objekt ersetzt werden soll. Zu beachten ist hierbei, dass aus Sicherheitsgründen der eingegebene Text automatisch in eckige Klammern gestellt wird (wird beispielsweise „Test“ im Feld eingegeben, so steht im Output „[Test]“). Im Bereich „Filter Language“ soll die Sprache eingestellt werden, in der die zu überprüfenden Informationen verfasst sind. D.h. die Sprache in der die eingehenden E-Mails

hauptsächlich verfasst sind. Falls sprachlich gemischte Mails zu erwarten sind, so ist als Classifier Sprache bevorzugt Englisch zu wählen, da dieser Classifier auch bei anderen Sprache gute Erkennungsleistungen bietet. Mehr zum Thema „Classifier“ finden Sie im Kapitel 4.2 Natural-Language-Processing (NLP). Im Bereich „Message Delay Time“ kann die minimale und die maximale Verzögerungszeit in Sekunden angegeben werden. Der Anonymous Remailer wählt für jede zu sendende Nachricht eine zufällige Verzögerungszeit zwischen diesen beiden Parametern um eine optimale Durchmischung der eingehenden Nachrichten zu ermöglichen. Dies erschwert Dritten, insbesondere wenn die Nachrichten verschlüsselt empfangen und gesendet werden, die Zuordnung von eingehenden zu ausgehenden Nachrichten. Um dieses Feature noch zu verstärken, kann im Bereich „Miscellaneous“ ausgewählt werden, ob an den Nachrichten Body ein „Lorem Ipsum“-Text (siehe http://de.wikipedia.org/wiki/Lorem_ipsum) zufälliger Länge angehängt werden soll (direkt sichtbar für den Empfänger). Alternativ kann auch ein beliebiger selbst erstellter Text angehängt werden. Zusätzlich kann auch eine zufällige Anzahl an künstlichen Headern hinzugefügt werden (indirekt sichtbar für den Empfänger). Im Anschluss an die Beschreibung der verschiedenen Tabs wird auf diese beiden Features noch näher eingegangen, da sie auf externen Dateien aufbauen. Die folgende Abbildung zeigt das „Anonymisation“ Tab:

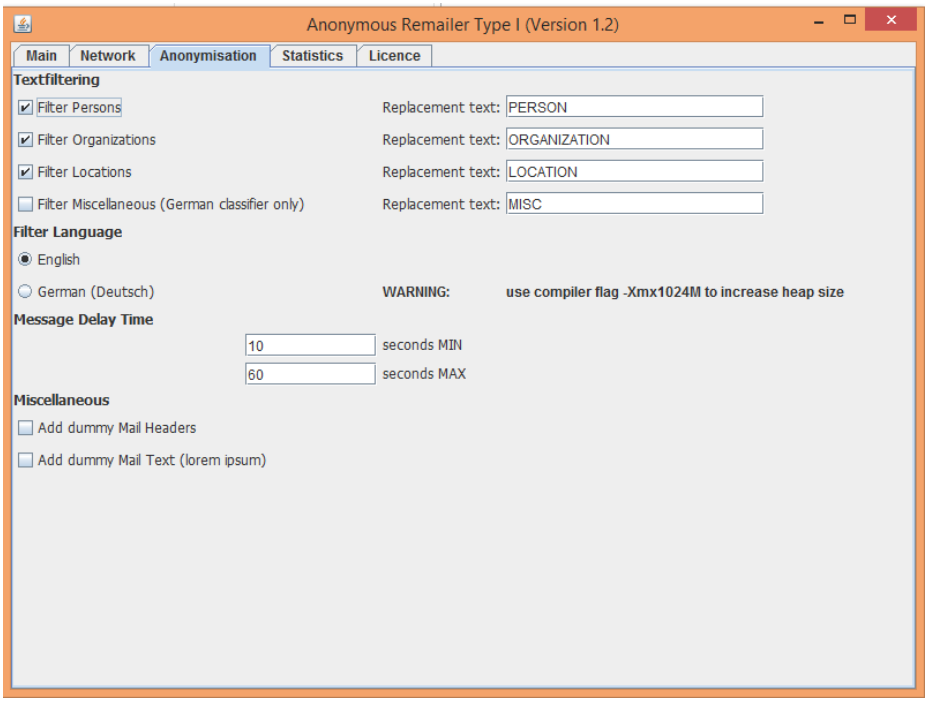

Abbildung 11: Anonymous Remailer „Anonymisation“ tab

Das „Statistics“ Tab wird in Echtzeit aktualisiert und zeigt die folgenden Werte an: „Received Emails“ zeigt die Anzahl der empfangenen Email Nachrichten an. Eine empfangene Nachricht kann an mehrere Empfänger weiter verschickt werden, wenn dies in der Nachricht so angegeben wurde. Für jeden Empfänger wird eine eigene Nachricht im Message Buffer angelegt, dies wird im Feld „Messages in Buffer“ angezeigt. Jede Nachricht im message Buffer

wird im Output Modul unabhängig voneinander behandelt/anonymisiert. Dies betrifft sowohl die „Add dummy Mail Headers“ als auch die „Add dummy Mail text“ Funktion. Nach Ablauf der festgelegten Verzögerungszeit wird die Nachricht an den Empfänger versandt. Daher verringert sich die Zahl im Feld „Messages in Buffer“ und erhöht sich im Feld „Send Emails“. Die folgende Abbildung zeigt das „Statistics“ Tab:

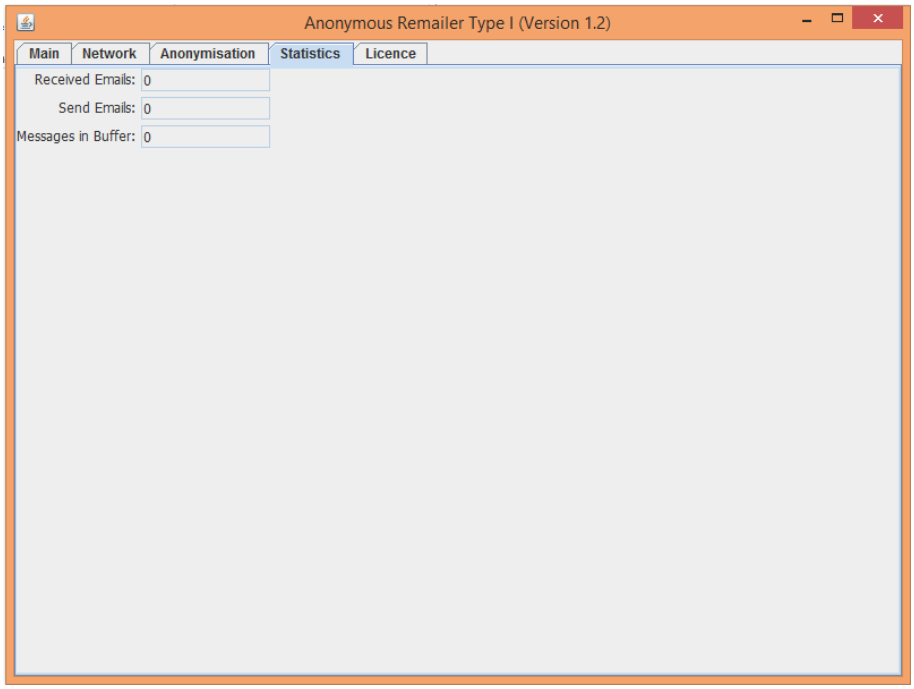

Abbildung 12: Anonymous Remailer „Statistics“ tab

Im „Licence“ Tab werden die eingesetzten Frameworks aufgelistet und deren Lizenzen. Die Lizenzen werden gemeinsam mit der Software im Ordner „licenses“ ausgeliefert. Die folgende Abbildung zeigt das „Licence“ Tab:

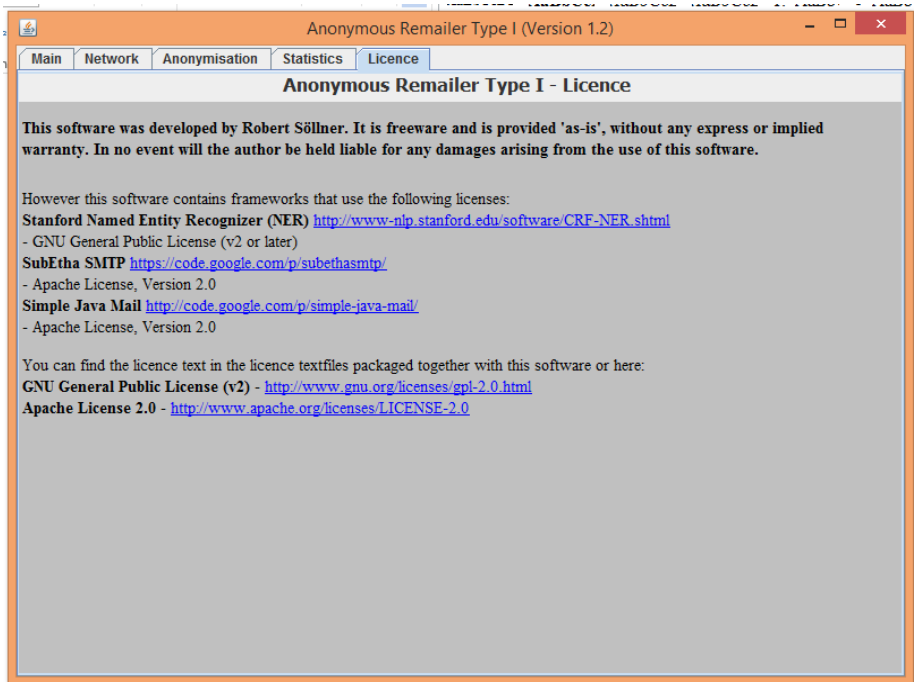

Abbildung 13: Anonymous Remailer „Licence“ tab

4.3.2. Ein- und Ausgabedateien

Eingabedateien

Bei Start des Anonymous Remailers werden die folgenden Dateien eingelesen:

```
inputfiles/dummyBodyText.txt
inputfiles/dummyHeaderText.txt
```

Diese Pfadnamen sind im Code vorgegeben und können nicht geändert werden.

dummyBodyText.txt

Die Datei enthält eine lange Folge von vorgeneriertem „Lorem ipsum“ Text. Z.B.:

```
Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy
eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam
voluptua. At vero eos …
```

Darunter versteht man eine pseudo-lateinische Sprache die dazu verwendet wird die typischen Wortlängen und Buchstaben-Verteilungen in westeuropäischen Sprachen zu simulieren. Im Gegensatz zu aktiv verwendeten Sprachen hat sie jedoch keine semantische Bedeutung und lenkt den Leser nicht vom eigentlichen Text der Nachricht ab.

Der Anonymous Remailer fügt jeder Nachricht im Anschluss an den „body“ Text eine zufällige Anzahl von Zeichen/Zeilen des „dummyBodyText.txt“ files an. Der Zeichenanfang wird ebenfalls bei jeder Nachricht zufällig gewählt. Der „Lorem ipsum“ Text kann – sofern gewünscht – gegen einen beliebigen, anderen Text ausgetauscht werden. Der gewählte Text sollte hinreichend lang und uniform sein, damit der Zufallsgenerator für jede Nachricht einen

pseudozufälligen Text auswählen kann. Ist der Text zu kurz oder stark unterschiedlich besteht die Gefahr der einfachen Mustererkennung.

dummyHeaderText.txt

Die Datei enthält vordefinierte Header, die den zu sendenden Nachrichten angehängt werden.. Diese sind nach der Struktur „Header-Element:“ *Leerzeichen* „Header-Wert“ *Neue Zeile* aufgebaut. Das Schema wird anhand einiger Beispiele demonstriert:

```
Comments: Some Comments
X-SpecialHeader: individual Message stated below
X-Company: MyCompany
X-Location: MyLocation
```

Der zu sendenden Nachricht wird eine zufällige Anzahl an Dummy Headern angehängt (zufällig bedeutet: kein, ein, mehrere oder alle Header im dummyHeaderText.txt file.

4.4. Benutzung: Nachrichten-Anonymisierung

Die Benutzung des Remailers für den Endanwender ist sehr einfach gestaltet. Der Benutzer muss die Internetadresse und den Port des Remailers kennen. Prinzipiell sind dies die Informationen, die auch für jeden anderen klassischen Mailservice benötigt werden. Der Benutzer kann nun den Server als Postausgangsserver in einem klassischen Mailclient angeben und ein E-Mail schreiben. Alternativ könnte der Benutzer auch mit einem Fernzugriffprotokoll z.B. telnet oder SSH sich mit dem Anonymous Remailer Server verbinden und das E-Mail über die Konsole eingeben, sofern dies vom Serveradministrator erlaubt wird. Dies hat den Vorteil, dass kein eigener Mailclient erforderlich ist. Unabhängig von der gewählten Methode muss bei der Nachrichteneingabe der vorgeschriebene Nachrichtenaufbau beachtet werden. Dieser geforderte Nachrichtenaufbau wird im Folgenden kurz beschrieben:

In der ersten Zeile des Body-Texts (nach dem Header) ist der Text „Anon-To:“ einzugeben. Gefolgt von einer mittels „;“ (Strichkomma) separierten Liste von Empfänger E-Mail Adressen. Möchte man selbst eine (eigenständig anonymisierte) Kopie der anonymisierten Nachricht erhalten, dann muss man die eigene E-Mail Adresse ebenfalls in der Liste der gewünschten Empfänger mit angegeben. Nach der Auflistung muss verpflichtend eine Leerzeile (blank line) erfolgen. In der nächsten Zeile kann die eigentliche Nachricht des Senders starten. Diese Vorgehensweise ist notwendig, da in die eigentlich „TO:“ Adresszeile die E-Mail Adresse des Anonymous Remailers eingetragen werden muss. Würde man in die „TO:“ Adresszeile zusätzlich die Adressen derjenigen eingetragen, die eine anonymisierte Nachricht bekommen sollen, so würde das E-Mail am Anonymous-Remailer vorbei direkt an diejenigen Empfänger gesendet werden. Dies würde dem Ziel der anonymen Nutzung widersprechen.

Sendet der Benutzer eine Nachricht an den Remailer, welche nicht dieser Struktur entspricht, so sendet der Remailer automatisch ein Hilfe/Usage-Mail an den jeweiligen Absender zurück:

```
USAGE: Write a normal email to the anonymous remailer. Add "Anon-To:" in the first
line of the email. Then add your recipients with a ";" separated list. Afterwards
please add a blank line. In the next line you can start the normal text.
Example:
Anon-To: mymail@mydomain1.com; myothermail@domain2.com

Sample Text body

INFO: Every recipient receives a separately anonymized email and cannot see the
other recipients. If you wish to receive an anonymized copy then please add your
email address to the recipients.
```

4.4.1. Testfälle und Erkennungsergebnisse

Der Anonymous Remailer arbeitet sehr flexibel. Er kann auf jedem Computer, der Java Programme ausführen kann und an ein Netzwerk angebunden ist, gestartet und getestet werden. Im „Network“ Tab des Anonymous Remailers muss der TCP Port spezifiziert werden, der für

den Erhalt von eingehende E-Mail Nachrichten verwendet werden soll. Die folgenden Abbildungen veranschaulichen die für den Test/die Tests eingestellte Konfiguration des Anonymous Remailers:

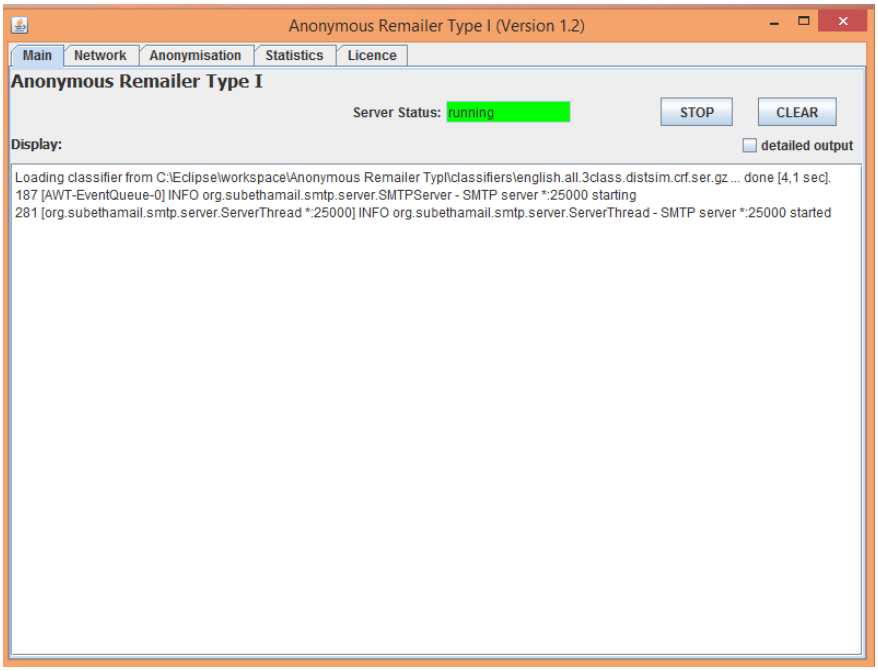

Abbildung 14: „Main“ Tab Konfiguration

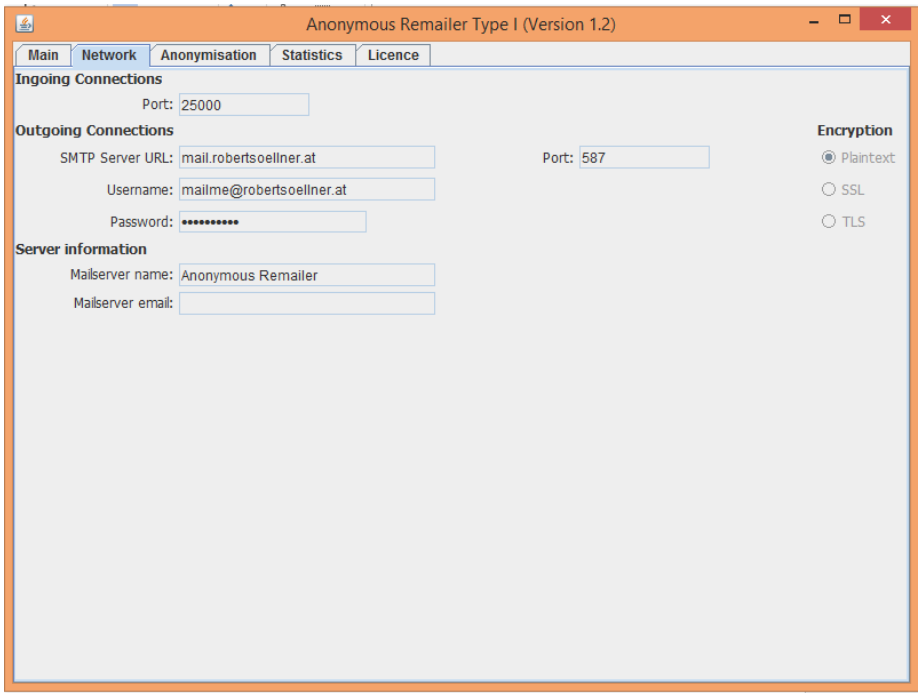

Abbildung 15: „Network“ Tab Konfiguration

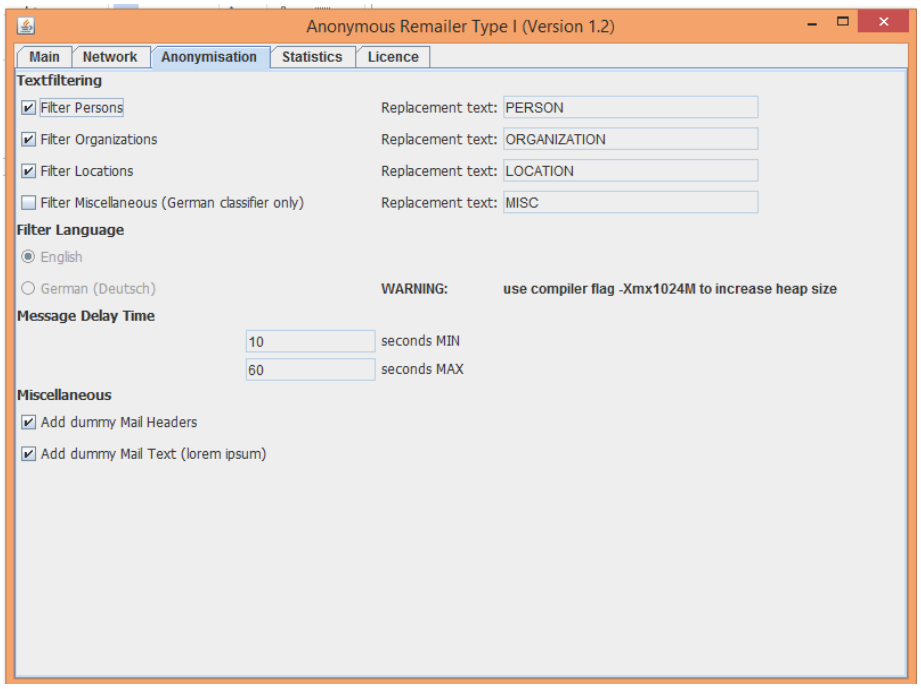

Abbildung 16: „Anonymisation“ Tab Konfiguration

Bei installiertem Telnet kann man sich direkt zum Anonymous Remailer verbinden: Dies funktioniert mittels des Befehls „telnet localhost 25000“ wenn der Anonymous Remailer am lokalen Computer läuft und für der SMTP Empfang für den Default Port „25000“ eingestellt wurde.

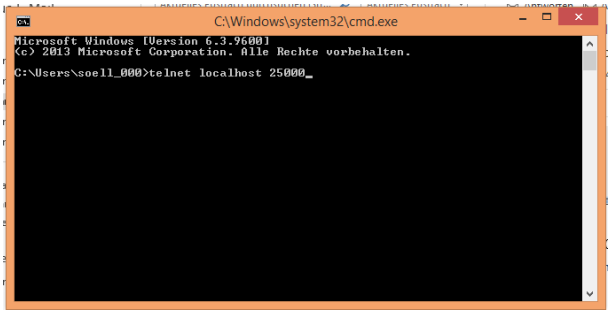

Abbildung 17: Shell Verbindungsaufbau zum Server

Protokoll der Testsession:

```
220 RoboNotebook ESMTP SubEthaSMTP null
EHLO localhost
250-RoboNotebook
250-8BITMIME
250 Ok
mail from: mailme@robertsoellner.at
250 Ok
rcpt to: uni@robertsoellner.at
250 Ok
data
354 End data with <CR><LF>.<CR><LF>
Subject: Test
Content-Type: text/plain;
```

```
Anon-To: uni@robertsoellner.at

Dear Mr. Bernstein,

This is a suitable text for testing the remailers anonymisation capability.

The Natural Language Processing Group at Stanford University is a team of faculty,
research scientists, postdocs, programmers and students who work together on
algorithms that allow computers to process and understand human languages. Stanford
is a center of competence in the topic of natural language processing within the
United States.

The NLP team members are:
- Chris Manning
statistical natural language processing, information extraction, text understanding
and text mining

- Dan Jurafsky
natural language understanding, conversational speech and dialog, natural language
processing for the behavioural and social sciences

- Percy Liang
semantic parsing, probabilistic models for natural language processing, machine
learning, program induction

Sincerely,
Robert Soellner
.
250 Ok
quit
221 closing channel

Quelle: http://nlp.stanford.edu/people.shtml
```

Tagged Anonymisation result:

```
Subject: Test

Dear Mr. <PERSON>Bernstein</PERSON>,

this is a suitable text for testing the remailers anonymisation capability. The
Natural Language Processing Group at <ORGANIZATION>Stanford
University</ORGANIZATION> is a team of faculty, research scientists, postdocs,
programmers and students who work together on algorithms that allow computers to
process and understand human languages. <ORGANIZATION>Stanford</ORGANIZATION> is a
center of competence in the topic of natural language processing within the
<LOCATION>United States</LOCATION>.

The <ORGANIZATION>NLP</ORGANIZATION> team members are:
- <PERSON>Chris Manning</PERSON>
statistical natural language processing, information extraction, text understanding
and text mining

- <PERSON>Dan Jurafsky</PERSON>
natural language understanding, conversational speech and dialog, natural language
processing for the behavioral and social sciences

- <PERSON>Percy Liang</PERSON>
semantic parsing, probabilistic models for natural language processing, machine
learning, program induction

Sincerely,
<PERSON>Robert Soellner</PERSON>
```

Quelle: http://nlp.stanford.edu/people.shtml

Anonymisierte Nachricht im Posteingang:

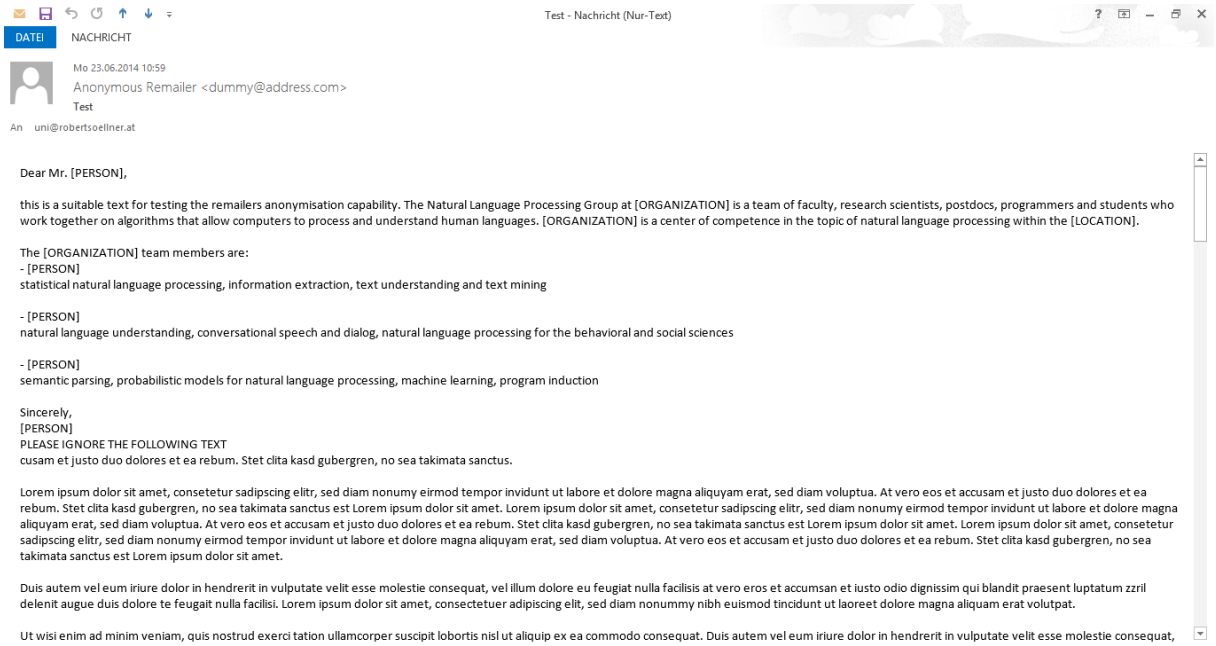

Abbildung 18: Ergebnis: Anonymisiertes E-Mail mit ersetzten Tags

Wie sich gut erkennen lässt wurden alle Tags korrekt gesetzt und die entsprechenend Kategorien anonymisiert. Die ausgezeichnete Erkennungsrate spricht für den von der Standford Universität entwickelten Named Entity Recognizer.

Es folgen die Header der Nachricht. Dort sind die vom Anonymous Remailer zufällig hinzugefügten künstlichen Header am Ende gut erkennbar:

```
X-Spam-Checker-Version: SpamAssassin 3.2.5 (2008-06-10) on www8.webmachine.eu
X-Spam-Level:
X-Spam-Status: No, score=0.1 required=7.0 tests=AWL,MISSING_MID,RDNS_NONE
       autolearn=no version=3.2.5
Received: (qmail 1381 invoked from network); 23 Jun 2014 10:58:28 +0200
Received: from unknown (HELO RoboNotebook) (217.14.232.55)
  by www8.webmachine.eu with SMTP; 23 Jun 2014 10:58:27 +0200
Date: Mon, 23 Jun 2014 10:58:37 +0200 (CEST)
From: Anonymous Remailer <dummy@address.com>
To:  <uni@robertsoellner.at>
Subject: Test
Comments: Some Comments
MIME-Version: 1.0
Content-Type: multipart/mixed;
       boundary="----=_Part_3_94757850.1403513917095"
X-Company: MyCompany
X-Location: MyLocation
```

Wenn die Header realistisch gewählt wurden (es wir empfohlen offiziell verwendete Header Tags zu verwenden), dann ist es nicht möglich sie von „normalen“ (ursprünglichen) Header der Nachricht zu unterscheiden. Dies garantiert eine zuverlässsige Verschleierung der Originalgröße des Headers und hilft damit die Deanonymisierung zu erschweren/verhindern.

4.4.2. Fremdsprachiger Text in Classifiern

Die im Anonymous Remailer für das Natural Language Processing verwendeten Classifier versuchen mit Hilfe errechneter Wahrscheinlichkeiten den Eingabetext, den passenden Kategorien (z.B. Namen, Orte, Organisationen, etc.) zu zuordnen. Siehe dazu Kapitel 4.2.1 Stanford Named Entity Recognizer (NER). Da die Verteilung der Erkennungsfaktoren von der verwendeten Sprache abhängig ist, werden Classifier immer für eine Zielsprache „trainiert“. Dieser Umstand erfordert es, den jeweiligen Classifier nach der zu erwartenden Eingangssprache zu wählen. Der Anonymous Remailer enthält per default die Classifier für die Sprachen Englisch und Deutsch. Beispielsweise sollte ein englischsprachiger Classifier verwendet werden, wenn der Anonymous Remailer hauptsächlich von englisch-sprachigen Usern verwendet wird. Ebenso sollte der deutschsprachige Classifier verwendetet werden, wenn deutschsprachige Benutzter die Zielgruppe sind. Natürlich kann es vorkommen, dass auch ein deutschsprachiger Benutzer international kommunizieren möchte und daher ein englischsprachiges E-Mail schickt. Oder eine beliebige andere Sprache. Da sowohl Englisch als auch Deutsch der germanischen Sprachfamilie abstammen sind hier die „gegenseitigen“ Erkennungsleistungen ganz passabel. Ändert sich dies, wenn man die Nachricht in einer aus der romanischen Sprachfamilie abstammenden Sprache wie z.B. Französisch verfasst?

Test mit französischem E-Mail:

```
Bonjour Madame Chung,

Je dois travailler samedi à l'atelier de mon père, Monsieur Luce Baptiste, donc je
ne pourrai pas me joindre à vous pour finaliser le projet. L’atelier est à Paris.
Je vous envoie ma partie en espérant que c'est satisfaisant. Il manque une
information, mais j'ai envoyé un message hier soir en Nouvelle-Zélande. J'espère
qu'ils comprendront que j'ai besoin d'une réponse rapidement.
Je serai à la maison vers 18h00 samedi et aussi dimanche si vous avez des questions
à me poser.

À lundi!
Julien Lavoie
```

Sollte wie folgt kategorisiert werden:

```
Bonjour <PERSON>Madame Chung</PERSON>,

Je dois travailler samedi à l'atelier de mon père, <PERSON>Monsieur Luce
Baptiste</PERSON>, donc je ne pourrai pas me joindre à vous pour finaliser le
projet. L’atelier est à <LOCATION>Paris</LOCATION>. Je vous envoie ma partie en
espérant que c'est satisfaisant. Il manque une information, mais j'ai envoyé un
message hier soir en <LOCATION>Nouvelle-Zélande</LOCATION>. J'espère qu'ils
comprendront que j'ai besoin d'une réponse rapidement.
Je serai à la maison vers 18h00 samedi et aussi dimanche si vous avez des questions
à me poser.

À lundi!
<PERSON>Julien Lavoie</PERSON>
```

Testresultat:

<PERSON>Bonjour **Madame Chung</PERSON>**,

Je dois travailler samedi à l'atelier de mon père, **<PERSON>Monsieur Luce Baptiste</PERSON>**, donc je ne pourrai pas me joindre à vous pour finaliser le projet. L'atelier est à **<LOCATION>Paris</LOCATION>**. Je vous envoie ma partie en espérant que c'est satisfaisant. Il manque une information, mais j'ai envoyé un message hier soir en **Nouvelle-Zélande**. J'espère qu'ils comprendront que j'ai besoin d'une réponse rapidement. Je serai à la maison vers 18h00 samedi et aussi dimanche si vous avez des questions à me poser.

À lundi!
<PERSON>Julien Lavoie</PERSON>

Es wurden bis auf „Nouvelle-Zélande“ (Neuseeland) und das Prefix „Bonjour“ (Guten Tag) alle Tags richtig gesetzt. Man kann daher darauf schließen, dass auch in fremdsprachigen Texten, Personennamen oder namensgleiche Orte einigermaßen zuverlässig erkannt werden. Bei exotischeren Benennungen wie z.B. „Nouvelle-Zélande“ (Neuseeland) versagt die Erkennung jedoch. Da die Namensfilterung sicher die Hauptaufgabe des Anonymous Remailers ist, kann die Aussage getroffen werden, dass der englische Classifier ebenfalls gut mit anderen Sprachen umgehen kann (sofern zumindest das lateinische Zeichensystem verwendet wird). Wenn eine möglichst exakte Erkennung gewünscht ist und dies auch bei anderen Kategorien, so muss ein für die jeweilige Sprache „trainierter“ Classifier eingesetzt werden.

4.4.3. Statistische Auswertungen

Die genaue Erkennungsleistung der beiden vorinstallierten Classifier (Englisch und Deutsch) ist anhand von Einzelbeispielen nicht ohne weiteres erkennbar, da sie stark vom zu klassifizierenden Testsample abhängig ist. Um die statistische Zuverlässigkeit zu ermitteln wurden für den deutschsprachigen Classifier fünf und für den englischsprachigen Classifier zehn Testsampels mit jeweils einer vielzahl an Tags an den Anonymous Remailer geschickt und anschließend die Fehlerrate ermittelt. Es werden dabei alle korrekt klassifizierenden Elemente gezählt (correct item = r). Ebenso werden nicht-erfolgte Klassifizierungen (missing item = m) und zuviel-erfolgte Klassifizierungen (false positive item = f) gezählt. Dies ergibt eine statistische Auswertung der Erkennungsrate über diverse Testsamples.

Im folgenden werden einige Textbeispiele vorgestellt und jeweils das Klassifierungsergebnis und die Auswertung demonstriert. Eine vollständige Auflistung der Testnachrichten sowie deren Klassifizierungsergebnis und die statistische Einzelauswertung finden Sie im Anhang im Appendix A (Classified Text).

Textsample:

Classifier DEUTSCH, Beispielnachricht 1

Lieber Peter,

Unter diesen Dokumenten ist auch die komplette Liste der von der NATO geplanten und durchgeführten "gezielten Tötungen". Im Militärjargon ist es die "Joint Prioritized Effect List" (JPEL). Diese Liste belegt, dass die Strategie der gezielten Tötungen nicht bloß als letztmögliches Mittel zur Bekämpfung der Taliban eingesetzt wurde, sondern Teil des regulären Vorgehens in diesem Krieg war.

Der "Spiegel" konnte Dokumente aus den Jahren 2009 bis 2011 einsehen, zu dieser Zeit war Stanley McChrystal ISAF-Kommandeur, der im Juni 2010 von David Petraeus abgelöst wurde. Schon 2009 schickte US-Präsident Barack Obama 33.000 zusätzliche Soldaten nach Afghanistan. Eine der blutigsten Phasen des Krieges begann. 2009 starben 2.412 Zivilisten, für ein Viertel dieser Toten waren NATO-Truppen und afghanische Sicherheitskräfte verantwortlich. Auch die Zahl der Einsätze gegen die Taliban nahm zu.

Hinter der Eskalation stand die Strategie von Petraeus: Er wollte die Gegner zuerst mittels einer "Säuberungsphase" schwächen, um anschließend das entstandene Machtvakuum mit lokalen Kräften zu füllen. Danach sollte eine Phase der Stabilisierung folgen.

Mfg
Robert Söllner
(Quelle: http://derstandard.at/2000009887439/NATO-Einsatz-in-Afghanistan-Gezielte-Toetungen-waren-Alltag)

Klassifizierungsergebnis:

Lieber **<I-PER>Peter</I-PER>**,

Unter diesen Dokumenten ist auch die komplette Liste der von der **<I-ORG>NATO</I-ORG>** geplanten und durchgeführten "gezielten Tötungen". Im Militärjargon ist es die **<I-ORG>"Joint Prioritized Effect List</I-ORG>" (<I-ORG>JPEL</I-ORG>)**. Diese Liste belegt, dass die Strategie der gezielten Tötungen nicht bloß als letztmögliches Mittel zur Bekämpfung der Taliban eingesetzt wurde, sondern Teil des regulären Vorgehens in diesem Krieg war.

Der **"Spiegel"** konnte Dokumente aus den Jahren 2009 bis 2011 einsehen, zu dieser Zeit war **<I-PER>Stanley McChrystal</I-PER>** ***ISAF***-Kommandeur, der im Juni 2010 von **<I-PER>David Petraeus</I-PER>** abgelöst wurde. Schon 2009 schickte US-Präsident ***Barack Obama*** 33.000 zusätzliche Soldaten nach **<I-LOC>Afghanistan</I-LOC>**. Eine der blutigsten Phasen des Krieges begann. 2009 starben 2.412 Zivilisten, für ein Viertel dieser Toten waren **<I-MISC>NATO-Truppen</I-MISC>** und **<I-MISC>afghanische</I-MISC>** Sicherheitskräfte verantwortlich. Auch die Zahl der Einsätze gegen die Taliban nahm zu.

Hinter der Eskalation stand die Strategie von **<I-PER>Petraeus</I-PER>:** Er wollte die Gegner zuerst mittels einer "Säuberungsphase" schwächen, um anschließend das entstandene Machtvakuum mit lokalen Kräften zu füllen. Danach sollte eine Phase der Stabilisierung folgen.

Mfg
<I-PER>Robert Söllner</I-PER>
(Quelle: http://derstandard.at/2000009887439/NATO-Einsatz-in-Afghanistan-Gezielte-Toetungen-waren-Alltag)

Auswertung:

Personen (r/f/m): 5/0/1 , Organisationen (r/f/m): 2/2/2 , Orte (r/f/m): 2/0/0

Classifier ENGLISCH, Beispielnachricht 5

If lawmakers are to break out of the partisan cycle, Mr. Kingston said, they need to avoid being inundated by their constituents in an increasingly digital world where members of Congress find themselves under immediate pressure as events unfold.

"If new members allow their base to control their behavior up here they are going to be miserable," said Mr. Kingston, who has seen the rising influence of Tea Party activists in Houston on Republican lawmakers. "While the voters might be yelling and screaming at you to do something, that's not your job.

"You have to look at all the information and then make the best determination as to what's going to be best for America," he said. "Sometimes you have to have disagreements with your own party along the way, and that is O.K."

A similar sentiment was expressed by Mr. Harkin, who was the principal sponsor of the Americans With Disabilities Act of 1990, and 18 years later was the chief sponsor of a law that expanded disability rights again by overturning several FBI decisions.

(Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html?_r=0)

Klassifizierungsergebnis:

If lawmakers are to break out of the partisan cycle, Mr. **<PERSON>Kingston</PERSON>** said, they need to avoid being inundated by their constituents in an increasingly digital world where members of **<ORGANIZATION>Congress</ORGANIZATION>** find themselves under immediate pressure as events unfold.

If new members allow their base to control their behavior up here they are going to be miserable, said Mr. **<PERSON>Kingston</PERSON>**, who has seen the rising influence of **<ORGANIZATION>Tea Party</ORGANIZATION>** activists in **<LOCATION>Houston</LOCATION>** on **<ORGANIZATION>Republican</ORGANIZATION>** lawmakers. While the voters might be yelling and screaming at you to do something, thats not your job.

You have to look at all the information and then make the best determination as to whats going to be best for **<LOCATION>America</LOCATION>**, he said. Sometimes you have to have disagreements with your own party along the way, and that is O.K.

A similar sentiment was expressed by Mr. **<PERSON>Harkin</PERSON>**, who was the principal sponsor of the Americans With Disabilities Act of 1990, and 18 years later was the chief sponsor of a law that expanded disability rights again by overturning several **<ORGANIZATION>FBI</ORGANIZATION>** decisions.

(Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html?_r=0)

Auswertung:

Person (r/f/m): 3/0/0 , Organization (r/f/m): 4/0/0 , Location (r/f/m): 2/0/0

Die Einzelauswertung der Textsamples spiegelt sehr gut die jweiligen Stärken und Schwächen der Classifier wieder. Bei der Hauptaufgabe der Classifier, der Namenserkennung und –filterung erkennt der deutsche Classifier das ein oder andere Tag nicht. Der Englische Classifier hingegeben arbeitet extrem zuverlässig und leistet sich keinen einzigen Fehler. Um dies zu veranschaulichen sind im Folgenden die Detailergebnisse der statitistischen Auswertung jeweils für den deutschen und anschließend für den englischen Classifier gelistet:

Auswertung Detailergebnisse:

DE:

1) Personen (r/f/m): 5/0/1 , Organisationen (r/f/m): 2/2/2 , Orte (r/f/m): 2/0/0

2) Personen (r/f/m): 4/1/3 , Organisationen (r/f/m): 0/0/1 , Orte (r/f/m): 1/0/0
3) Personen (r/f/m): 4/0/1 , Organisationen (r/f/m): 0/0/0 , Orte (r/f/m): 7/0/1
4) Personen (r/f/m): 5/0/0 , Organisationen (r/f/m): 4/0/0 , Orte (r/f/m): 0/0/1
5) Personen (r/f/m): 4/0/3 , Organisationen (r/f/m): 6/0/1 , Orte (r/f/m): 3/0/1

Summe:

Personen (r/f/m): 22/1/8 , Organisationen (r/f/m): 12/2/4 , Orte (r/f/m): 13/0/3

EN:

1) Person (r/f/m): 4/0/0 , Organization (r/f/m): 1/0/0 , Location (r/f/m): 5/0/0
2) Person (r/f/m): 4/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 1/0/0
3) Person (r/f/m): 3/0/0 , Organization (r/f/m): 3/0/0 , Location (r/f/m): 3/0/0
4) Person (r/f/m): 1/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 6/0/0
5) Person (r/f/m): 3/0/0 , Organization (r/f/m): 4/0/0 , Location (r/f/m): 2/0/0
6) Person (r/f/m): 6/0/1 , Organization (r/f/m): 1/0/0 , Location (r/f/m): 7/0/0
7) Person (r/f/m): 5/0/1 , Organization (r/f/m): 6/0/1 , Location (r/f/m): 0/0/0
8) Person (r/f/m): 8/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 3/0/0
9) Person (r/f/m): 12/0/0 , Organization (r/f/m): 3/0/0 , Location (r/f/m): 5/0/1
10) Person (r/f/m): 13/0/1 , Organization (r/f/m): 1/1/0 , Location (r/f/m): 4/0/0

Summe:

Person (r/f/m): 59/0/3 , Organization (r/f/m): 25/1/1 , Location (r/f/m): 36/0/1

Anhand der angegeben Summenwerte lässt sich die Genauigkeit der Classifier tabellarisch und prozentuell angeben. Es ist dabei zu beachten, dass die angegebenen prozentuellen Erkennungsraten auf der Datengrundlage des Testsamples basieren. Wenn die Datengrundlage aufgrund anderer Testsamples verändert wird, so können sich auch die prozentuellen Werte ändern. Die folgende Tabelle sollte daher als tabellarische Aufbereitung der Testergbenisse der Masterarbeit gewertet werden:

Used Classifier	Total no. of items	Correct items	False positive	Missing items	**% Classified**	**% False pos**
English	125	120	1	5	**96%**	**0,8%**
Deutsch	65	47	3	15	**72,31%**	**4,62%**

Berechnung:

% Classified = Correct no. of items / (Total no. of items / 100)
% False positive = False pos. no. of items / (Total no. of items / 100)

Anmerkung zur statistischen Auswertung der Classifier Ergebnisse:

DE

Wörter wie z.B. Der „Spiegel“, "Die Welt" wurden als Fehler gewertet, da sie im Textkontext als Organisationen gelten. Vom Classifier wurden sie jedoch nicht erkannt, was vermutlich der Tatsache geschulted ist, dass die Wörten im normalen Sprachgebrauch ebenfalls vorkommen und daher nicht klar als Organisation erkannt werden können. Ebenso wurde gewertet, dass "Joint Prioritized Effect List" (JPEL) jeweils fälschlich als Organisation erkannt wurde da dies eine Liste und keine Organsiation ist. Bereits an diesen Beispielen kann man gut erkennen, wie komplex die Verarbeitung natürlicher Sprache ist. Je nach Kontext haben Wörter unterschiedliche Bedeutungen und würden unterschiedlichen Kategorien zugeordnet werden. Diese kontextabhängige Klassifizierung ist äußerst schwierig und ein aktuelles Forschungsfeld im Bereich des „Natural Language Processings“.

Da der deutsche Classifier ein 4class-Classifier ist und die vierte Kategorie „Miscelanious“ nicht näher definiert ist, wurden Erkennungen in dieser Kategorie je nach Art entweder der Kategorie „Ort“ oder der Kategorie „Organisation“ zugerechnet, wenn dies korrekt war. Erfolgte eine falsche Klassifizierung so wurde diese in der jeweiligen passenden Kategorie als zusätzlicher Fehler mit eingerechnet. Daher wurde die vierte Klasse in der Auswertung nicht extra angeführt um den Vergleich mit dem englischen 3class-Classifier zu ermöglichen.

4.4.4. Zusammenfassung der Testergebnisse

Insbesondere der englische Classifier weist eine hervorragende Erkennungsleistung von 96% und sehr geringe False Positive Trefferergebnisse von 0,8% auf. Man würde eine wesentliche höhere Anzahl von Testsamples benötigen um eine statistisch, korrekte Wahrscheinlichkeit zu erhalten. In dieser Masterarbeit wäre beispielsweise nach 5 Testsamples die Erkennungsleistung bei 100% gelegen. Im Praxistest ist eine 100% Erkennungssleistung fast nicht zu erreichen, da unsere Sprache oft nicht eindeutig ist. Es erfordert Fachwissen oder Kontextwissen des Lesers um korrekt „verstanden“ zu werden. Da dem Classifier in der Regel dieses Wissen fehlt, kann es zu Klassifizierungsfehlern kommen. Desweiteren hat der Fremsprachentest ergeben, dass auch bei Fremdsprachen der englischsprachige Classifier akzeptable Ergebnisse liefert. Er ist daher für den zu erwartenden Nachrichtenverkehr im Anonymous Remailer sehr gut geeignet. Sollte der Anonymous Remailer ausschließlich im deutschsprachigen Bereich verwendet werden, so ist der Einsatz der deutschsprachigen Classifier Variante zu empfehlen.

Generell ist zu betonen, dass die Erkennungsleistung die wichtigste Kennzahl für die effektive Bewertung von Classifiern ist. Dies liegt daran dass bei der Anonymisierung bereits ein einziges nicht erkanntes Wort (wenn es sich beispielsweise um ein prägnantes Schlüsselwort handelt) ausreichend sein kann um die gesamte durchgeführte Anonymisierung eines Textes

zu kompromitieren. Daher muss es oberste Maxime eines Classifiers sein danach zu streben eine vollständige Erkennungsleistung zu erreichen. Da eine hohe Erkennungsleistung auch durch ensprechend viele fälschlich erkannte Wörter (= false positive rate) erreicht werden könnte ist auch diese Kennzahl für die qualitative Bewertung eines Classifiers relevant. Je gringer die „false positive rate“ eines Classifiers ist, desto weniger Informationen gehen für den Leser eines Textes verloren. Eine hohe „false positive rate“ kann dazu führen, dass für den Leser relevante Informationen verfälscht oder verloren gehen und dadurch der Text nicht mehr die beabsichtigten Informationen an den Leser vermittelt

5. IT-Sicherheit

5.1. Einleitung

Das Kapitel IT-Sicherheit schafft Bewusstsein für die Notwendigkeit von Sicherheit auch in der elektronischen Welt. Es erklärt warum Sicherheit in der IT Branche früher weniger bedeutend war als heutzutage und nennt Schutzziele anhand derer der Begriff IT-Sicherheit definiert werden kann. Vor allem schafft es Bewusstsein dafür, dass es im Normalfall keine 100% Sicherheit gibt. Sicherheit ist relativ und die Gewichtung der Schutzziele hängt u.a. von den Anforderungen für ein System und dem Anwendungszweck ab. Des Weiteren wird auf generelle Bedrohungen und Begriffe der IT-Sicherheit eingegangen und es wird ein generisches Schadensmodell vorgestellt. Es wird ein besonderer Fokus auf das Schutzziel der Anonymisierung und Pseudomisierung gelegt und das Spannungsfeld mit den anderen Schutzzielen erläutert.

IT-Systeme gewinnen stetig an Bedeutung und sind in vielen Bereichen sowohl im privaten wie auch im beruflichen Umfeld nicht mehr wegzudenken. IT steht für Informationstechnologie, dies verdeutlicht, dass IT-Systeme als Informationsquellen dienen und zur Speicherung von Informationen genutzt werden. Ein bekannter Spruch lautet: „Wissen ist Macht“. Der Fall „Wikileaks“ zeigte welche Brisanz die Veröffentlichung von Informationen, die eigentlich nicht für die Öffentlichkeit bestimmt waren, haben kann. Unternehmen sind stark an der Wahrung ihrer Geschäftsgeheimnisse interessiert, ebenso Privatpersonen, die z.B. ihr Onlinebanking Passwort mit Sicherheit nicht der Öffentlichkeit bekannt geben würden. Der Zugriff auf Informationen soll nur Berechtigten möglich sein. IT-Sicherheit beschreibt u.a. die Maßnahmen zum Schutz von Informationen in IT-Systemen.

Der Begriff IT-Sicherheit wird daher zumeist analog mit dem Begriff Informationssicherheit verwendet. Davon zu unterscheiden ist der Datenschutz (engl. Privacy). Datenschutz beschreibt Maßnahmen zum sicheren Umgang mit personenbezogenen Informationen in der Datenverarbeitung und bei der Weitergabe an Dritte. Datenschutz steht für die Idee, dass jeder Mensch grundsätzlich selbst entscheiden kann, wem wann welche seiner persönlichen Daten zugänglich sein sollen und hat zum Ziel den so genannten „gläsernen Menschen“ zu verhindern. Für den Bereich Datenschutz existieren zahlreiche staatliche und überstaatliche Reglementierungen beispielsweise die EU-Richtlinie 95/46/EG (Datenschutzrichtlinie) sowie die EU-Richtlinie 2002/58/EG (Datenschutzrichtlinie für elektronische Kommunikation). Gesetzliche Rahmenbedingungen spielen auch in der IT-Sicherheit eine nicht zu unterschätzende Rolle, diese Masterarbeit beschäftigt sich jedoch Großteils mit der technischen Sicherheit.

Geschichtliche Entwicklung von IT-Sicherheit

IT-Sicherheit ist der laufenden, technischen und gesellschaftlichen Weiterentwicklung unterworfen. Seit vielen Jahrzehnten wird über die Sicherheit und Zuverlässigkeit von IT-Systemen nachgedacht und daran geforscht. Dabei stiegen die Anforderungen an sichere Systeme mit der technischen Leistungsfähigkeit und den Anwendungsbereichen der Systeme. Im Laufe der Zeit haben sich Schutzziele entwickelt, die generelle Anforderungen an die IT-Sicherheit definieren. „Vor dreißig Jahren wurde Sicherheit nahezu mit Vertraulichkeit gleichgesetzt, beispielsweise im Orange Book [DOD85]. Vor 25 Jahren wurde Integrität der Information und Verfügbarkeit der Funktionalität hinzugefügt, beispielsweise von Voydock and Kent [VOK83] und in den europäischen Kriterien für Sicherheitsevaluationen [OPE91]. Vor zwanzig Jahren kam die Zurechenbarkeit als viertes Schutzziel hinzu, beispielsweise in den Kanadischen Kriterien [GOC92]“ [BEA10].

Außerhalb des Hauptstroms der staatlich dominierten Sicherheitsforschung wurden Anonymität und Unbeobachtbarkeit ebenso relevante Themen, als der Fortschritt der Speichertechnologie soweit war, dass alle personenbezogenen Daten nahezu kostenlos und unbegrenzt gespeichert werden konnten. In den vergangenen Jahren förderten Versuche mancher Regierungen, den Gebrauch der Kryptographie zu reglementieren, und das Bestreben der Musik- und Filmindustrie, Techniken zur Kontrolle des Kopierens und Verbreitens digitaler Inhalte zu entwickeln, ungemein die Entwicklung der Steganographie. Dies ist die alte Kunst und entstehende Wissenschaft, wie Information in anderen, unverfänglichen Daten versteckt werden kann (Verdecktheit). Mobilfunknetze, die es ermöglichen, Personen unabhängig davon, wo sie sich befinden und was sie gerade tun, zu erreichen, brachte das Schutzziel Erreichbarkeit ins Bewusstsein. Damit ist die Kontrollmöglichkeit gemeint, wer wen unter welchen Umständen mittels welcher Medien und Kommunikationsdienste erreichen kann. Elektronischer Handel schärfte den Sinn für Verbindlichkeit/Zurechenbarkeit, d.h. Teilnehmer müssen ihre rechtlichen Pflichten innerhalb einer vernünftigen Zeitspanne erfüllen.

Spannungsfeld Sicherheitsanforderungen

Generell kann man feststellen, dass die von unterschiedlichen Beteiligten erwarteten Sicherheitsanforderungen üblicherweise sehr verschieden sind. Im klassischen zentralisierten Rechenzentrum spielten und spielen Fragen der Verfügbarkeit, Ausfallsicherheit, und Fehlertoleranz von Systemen und Daten eine große Rolle. Der Trend zu verteilten Systemen führt hingegen zu einer stärkeren Beachtung der Kommunikationssicherheit und Integrität, d. h. der Sicherung von Daten z. B. gegen Verfälschung, Manipulation und unbefugtes Mitlesen während der Übertragung. In komplexen IT-Systemen, wie es unsere heutigen und zukünftigen Kommunikationsnetze sind, handeln verschiedene Subjekte (Organisationen, Personen). Sie können dabei nicht nur kommunizieren und kooperieren, sondern auch konkurrieren (z. B. um Betriebsmittel), sabotieren (z. B. Kommunikation behindern, stören, blockieren), fingieren (z. B. Identitäten vortäuschen, Daten verändern) oder abhören (z. B. bespitzeln, lauschen) und vieles mehr. [FEP02]

Um Funktion und Eigenschaften eines Systems beim Eintreten von nicht erwünschten Ereignissen aufrecht zu erhalten, sind daher Schutzmaßnahmen erforderlich. IT-Systeme (einschließlich der Übertragungsstrecken) müssen dazu gegen unbeabsichtigte Fehler und Ereignisse (z. B. höhere Gewalt, technische Fehler, Fahrlässigkeit, Programmierfehler, Verschleiß, Havarien) und beabsichtigte Angriffe (z. B. Abhören, Manipulation und Zerstören von Informationen, aber auch von Software und Hardware) von außen (z. B. Hacker oder Terroristen mit Sprengstoff) und innen (z. B. Administratoren, Programmierer) gesichert werden. [FEP02]

5.2. Schutzziele

Ein IT-System muss die Funktionssicherheit (engl. safety) gewährleisten, d.h. es soll unter allen normalen Betriebsbedingungen funktionieren und zuverlässig sein. Funktionssicherheit stellt die Grundlage für die Informationssicherheit (engl. security) und Datensicherheit (engl. protection) dar. Security ist der Schutz (inkl. aller Maßnahmen, diesen Schutz zu erreichen) des IT-Systems vor bedrohendem Verhalten der Systemumgebung, d.h. der Schutz vor beabsichtigten (böswilligen) Verhalten. Safety wird für den Schutz der Rechnerumgebung vor unbeabsichtigten Ereignissen, d.h. vor allem menschliches oder technisches Versagen, verwendet, d.h. den Schutz der Systemumgebung (z.B. von Personen) von Fehlverhalten des Systems [DIN EN 1999, ISO/IEC TR 13335.1 1996, ISO/IEC 2003]. Generell unterscheidet man zwischen passiven und aktiven Angreifern. Während passive Angreifer nur beobachten bzw. abhören (d.h. die Kommunikation nicht stören oder verändern – unautorisierte Informationsgewinnung), üben aktive Angreifer einen unerwünschten Einfluss auf die Kommunikation aus (unautorisierte Modifikation, verletzt Integrität und/oder Verfügbarkeit eines IT-Systems). Daraus ergeben sich die verschiedenen Schutzziele, die diese unerwünschte Einflussnahme verhindern oder falls dies nicht möglich ist, zumindest aufdecken sollen. In der Praxis verwenden Angreifer sowohl passive als auch aktive Angriffstechniken. Beispielsweise werden zuerst passiv Informationen in einem Netzwerk gesammelt (z.B. verwendete Protokolle, Netzwerkstruktur, Zugriffscodes, etc.) um mit diesen Informationen dann aktiv ins Netzwerk einzugreifen. [POG07]

5.2.1. CIA-Modell

Im Laufe der Zeit entwickelten sich zahlreiche Schutzziele. In der Fachwelt wird des Öfteren darüber diskutiert, welches die Kernprinzipien der Informationssicherheit sind und ob diese erweitert werden sollten. Prinzipiell lassen sich alle anderen Schutzziele unter diese Kernprinzipien einordnen. Seit nun über zwanzig Jahren werden Vertraulichkeit (engl. Confidentiality), Integrität (engl. Integrity) und Verfügbarkeit (engl. Availability) als die drei Kernprinzipien der Informationssicherheit angesehen, dies wird oft als die sogenannte CIA-Triad bezeichnet. Abbildung 19 zeigt das CIA-Informationssicherheitsmodell:

Quelle: http://en.wikipedia.org/wiki/File:CIAJMK1209.png

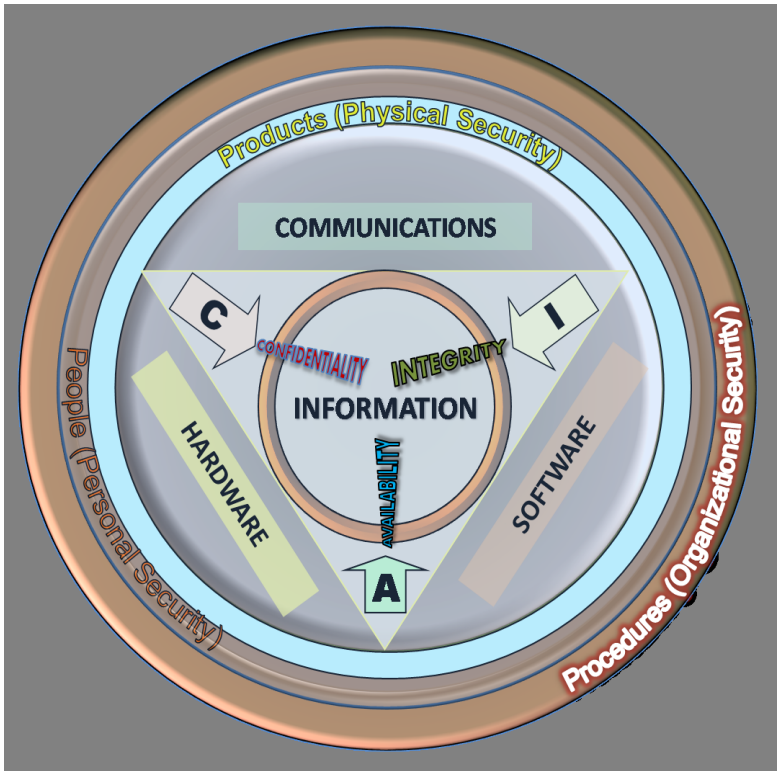

Abbildung 19: CIA-Informationssicherheitsmodell

Abbildung 19 zeigt die Kernkomponenten der Informationssicherheit. IT-Systeme setzen sich aus der eigentliche Hardware, der Software, die auf dieser Hardware läuft und der Kommunikation, die zwischen den Komponenten stattfindet, zusammen. IT-Sicherheit soll auf drei Ebenen wirken. Unternehmens-Policies und Sicherheitsvorschriften sollen sicherstellen, dass Personen (Administratoren, Benutzer, etc.), IT-Systeme unter Einhaltung der Schutzziele verwenden. Diese drei semantischen Dimensionen - Vertraulichkeit, Integrität und Verfügbarkeit - werden als Sicht der Verlässlichkeit bezeichnet, da sie vor allem auf die technische Sicherheit abzielen.

Aus Sicht der Verlässlichkeit gibt es drei Grundbedrohungen. Erstens unbefugter Informationsgewinn, das führt zur Beeinträchtigung oder zum Verlust der Vertraulichkeit von Ergebnisses oder Funktionen. Zweitens unbefugte Modifikation, dies beeinträchtigt oder kompromittiert die Integrität von Ergebnissen oder Funktionen. Drittens die unbefugte Veränderung der Funktionalität, dies beeinträchtigt oder verhindert die Verfügbarkeit des Funktionsablaufs oder der Ergebnisse. Ein sicheres IT-System muss daher im Stande sein die Vertraulichkeit, Integrität und Verfügbarkeit von Daten, Programmen und Geräten aufzubauen und bei jeder Nutzung zu erhalten. [DIE04]

Des Weiteren gibt es noch die Sicht der Beherrschbarkeit, die die Sicht der Betroffenen repräsentiert. Rechte oder schutzwürdige Belange von Personen dürfen durch das Vorhandensein oder die Nutzung von IT-Systemen nicht unzulässig beeinträchtigt werden. Bei elektronischen Geschäften (E-Business) ist es essentiell die Geschäftspartner zu identifizieren und die Echtheit der Dokumente (z.B. Rechnungen) sicher zu stellen, sowie den verbindlichen Charakter

geschäftlicher Zusagen sicher zu stellen (Verbindlichkeit) um Betrug zu verhindern oder aufzudecken. [DIE04]

Im Folgenden werden Schutzziele definiert, um eine umfassende Definition des Begriffs IT-Sicherheit zu ermöglichen:

- **Sicht der Verlässlichkeit**
  Vertraulichkeit, Integrität, Verfügbarkeit
- **Sicht der Beherrschbarkeit**
  Zurechenbarkeit, (Rechts-)Verbindlichkeit
- **Datenschutzziele**
  Anonymisierung und Pseudomysierung

5.2.2. Vertraulichkeit (engl. confidentiality)

„Ein System gewährleistet die Informationsvertraulichkeit, wenn es keine unautorisierte Informationsgewinnung ermöglicht.“ [ECK06]

„Die Gewährleistung der Eigenschaft der Informationsvertraulichkeit erfordert in datensicheren Systemen die Festlegung von Berechtigungen und Kontrollen der Art, dass sichergestellt ist, dass Subjekte nicht unautorisiert Kenntnis von Informationen erlangen. In informationssicheren Systemen sind Maßnahmen zur Festlegung sowie zur Kontrolle zulässiger Informationsflüsse zwischen den Subjekten des Systems erforderlich.“ [ECK06] Dadurch kann ausgeschlossen werden, dass Information zu unautorisierten Subjekten „durchsickert“. Das entsprechende Problem wird Confinement Problem genannt. Mit den Festlegungen zur Kontrolle der Informationsflüsse ist spezifiziert, welche Subjekte Kenntnis von welchen Informationen erlangen dürfen.

Anforderungen an die Informationsvertraulichkeit im weiteren Sinn werden durch Verschlüsselungstechniken erfüllt. Hierbei besteht das Ziel darin, die Daten geeignet zu transformieren, sodass unautorisierte Dritte nicht in der Lage sind die Daten sinnvoll zu interpretieren. Zur Gewährleistung von Vertraulichkeitsanforderungen sind somit spezielle Maßnahmen erforderlich, die über eine reine Kontrolle der Zugriffe auf die Datenobjekte hinausgehen. Verwendet werden hierbei, neben kryptografischen Verfahren, insbesondere auch Labeling-Techniken, wodurch Datenobjekte eine spezielle Sicherheitseinstufung erhalten. Durch spezielle Kontrollen kann sichergestellt werden, dass Information, die z.B. als sensitiv eingestuft ist, nicht in unautorisierte Hände gelangt. [ECK06]

5.2.3. Integrität (engl. integrity)

„Ein System gewährleistet die Datenintegrität, wenn es Subjekten nicht möglich ist, die zu schützenden Daten unautorisiert und unbemerkt zu manipulieren.“ [ECK06]

Die Eigenschaft der Datenintegrität erfordert zum einen die Festlegung von Rechten zur Nutzung von Daten. Beispiele hierfür sind Lese- oder Schreibberechtigungen für Dateien oder das Recht, von einem bestimmten Konto einen Betrag bis zu einer festgelegten Obergrenze abheben zu dürfen. Zum anderen sind Rechte an Subjekte zu vergeben, sodass diese autorisiert sind, die entsprechenden Zugriffsrechte wahrzunehmen. Abhängig von den damit getroffenen Festlegungen können Integritätsaussagen unterschiedlicher Qualität gemacht werden. So wird beispielsweise durch die Vergabe von Schreibberechtigungen an Dateien die Möglichkeit zur Modifikation des Datei-Objekts nicht weiter beschränkt, sodass Subjekte zu beliebigen Manipulationen berechtigt sind. Auf dieser Basis sind nur eingeschränkte Aussagen über die Integrität im Sinne einer authentischen, korrekten Funktionalität des Daten-Objekts möglich. Legt man demgegenüber die Berechtigungen zur Nutzung von Objekten durch wohldefinierte Methoden des Objekts fest, so werden die Nutzungsmöglichkeiten und damit die Manipulationsmöglichkeiten auf die Funktionalität dieser Zugriffsoperationen eingeschränkt. Die benötigten Mechanismen und Verfahren zur Gewährleistung des Schutzzieles der Datenintegrität gehören zum Bereich der Zugriffskontrolle. [ECK06]

Datenintegrität fordert, dass unautorisierte Manipulationen nicht unbemerkt bleiben dürfen. Das bedeutet, dass in Umgebungen, in denen eine solche Manipulation nicht a priori verhindert werden kann, Techniken erforderlich sind, mit deren Hilfe unautorisierte Manipulationen a posteriori erkennbar sind. Auf diese Weise kann verhindert werden, dass unautorisiert manipulierte Daten weiterverarbeitet werden und dadurch zumindest der mögliche Schaden begrenzt wird. Zu Erkennung von durchgeführten Datenveränderungen werden u.a. kryptografisch sichere Hashfunktionen eingesetzt. [ECK06]

5.2.4. Verfügbarkeit (engl. availability)

„Ein System gewährleistet Verfügbarkeit, wenn authentifizierte und autorisierte Subjekte in der Wahrnehmung ihrer Berechtigungen nicht unautorisiert beeinträchtigt werden können.“ [ECK06] Eine autorisierte Beeinträchtigung wäre z.B. das Prozess-Scheduling bei Betriebssystemen. Da es sich hierbei um „normale“ Managementmaßnahmen handelt, die ausgeführt werden müssen, stellt dies a priori noch keine Verletzung der Verfügbarkeit dar.

Die Verfügbarkeit setzt sich sowohl aus der Zuverlässigkeit als auch aus der Fehlertoleranz eines Systems zusammen. Die Zuverlässigkeit ist die Eigenschaft eines Systems, Zuverlässigkeitsforderungen während vorgegebenen Anwendungsbedingungen zu erfüllen [laut DIN 40041]. Die Fehlertoleranz ist die Eigenschaft eines Systems, spezifizierte Funktionen mit einer begrenzten Zahl fehlerhafter Subsysteme zu erfüllen. Weitere Eigenschaften, die Verfügbarkeit eines Systems betreffend, sind Robustheit und Wiederherstellbarkeit. Robustheit ist die Eigenschaft eines Systems, eine bestimmte Mindestmenge an Funktionen der Gesamtfunktionalität abzuwickeln. Wiederherstellbarkeit ist die Eigenschaft eines Systems, von fehlerhafter zu korrekter Leistungserbringung zu gelangen und die betroffenen Daten wiederzugewinnen. Je besser diese Werte sind, desto höher wird in der Regel die Verfügbarkeit sein. [POH04]

Die Zuverlässigkeit eines Systems wird quantitativ mittels der Verfügbarkeit (V) angegeben, diese ist laut DIN 40041 wie folgt definiert: Verfügbarkeit ist die Wahrscheinlichkeit, ein System zu einem geg. Zeitpunkt in einem funktionsfähigen Zustand anzutreffen. Zur Berechnung dieses Wertes wird u.a. die MTBF (Mean Time Between Failure) verwendet. Dieser Wert gibt die durchschnittliche Zeit an, während der ein System funktionsfähig ist, bis ein fehlerhafter Zustand auftritt. Die durchschnittliche Zeit, die benötigt wird um von einem fehlerhaften Zustand wieder in einen funktionsfähigen Zustand zu gelangen, wird als MTTR (Mean Time To Repair) angegeben. Wenn nach Auftreten eines Fehlers stets repariert wird, so ist die gesamte Einsatzdauer (=bisherige Missionsdauer) die Summe aus MTBF und MTTR. Für diesen in der Praxis üblichen Fall ist die Verfügbarkeit V=MTBF/(MTBF+MTTR). [SKO11]

5.2.5. Zuordenbarkeit (engl. accountability)

„Von jeder in einem IT-System ausgeführten Aktion (Vorgang, Prozess) muss während ihres Ablaufs und danach feststellbar sein, wem, d.h. welcher Instanz – insbesondere welcher Person – diese Aktion zuzuordnen ist. Es muss klar feststellbar sein, welches Subjekt sie ausgelöst und wer sie letztlich zu verantworten hat.“ [DIE04] Diese Forderung gilt nicht nur für die Aktionen selbst, sondern erst recht für deren Ergebnisse oder Auswirkungen.

In Computernetzwerken gibt es so gut wie keine Möglichkeit, sich von der Authentizität einer Person (z.B. eines Absenders oder Empfängers) oder eines Gegenstanden (z.B. eines Schrift- oder Musikstücks, einer Unterschrift, eines Bildes) unmittelbar und persönlich zu überzeugen. In Netzen ist immer mindestens eine transformierende und somit interpretierende Instanz (Hardware, Software, Gerät, Programm) zwischengeschaltet. Damit aber hängt die Antwort auf die Frage nach der Authentizität von Personen oder Gegenständen entscheidend davon ab, wie vertrauenswürdig diese „Übersetzer“ - die transformierenden Instanzen - wirklich sind, daher kommt der Authentifizierung große Bedeutung zu. Man unterscheidet „Authentisieren“ und „Authentifizieren“. Beide Begriffe sind auf das spätlat. authenticus (zuverlässig verbürgt; urschriftlich; eigenhändig) zurückzuführen, unterscheiden sich aber in ihrer Bedeutung. Während authentifizieren „beglaubigen; die Echtheit bezeugen“ heißt, bedeutet authentisieren „glaubwürdig, rechtsgültig machen“. Im EDV-Bereich authentisiert man sich etwa mit Hilfe eines Passworts (= Anbringung der Legitimation) und wird dann vom System authentifiziert (= Prüfung der Legitimation). Ein Subjekt behauptet eine Identität und diese muss verifiziert werden. Dies ist Teil des zweistufigen Verfahrens Identifizieren und Authentifizieren und wird als Authentifizierung (eng. authentification) bezeichnet. Nicht vernachlässigt werden darf die Sicht der Beherrschbarkeit (engl. governance). Darunter können die folgenden Aspekte eines Systems verstanden werden: Steuerung (engl. controlling), eindeutige Identifizierung (engl. identification) der Handlungsfolgen und die Zurechenbarkeit (engl. accountability) der ergriffenen Handlungen zu einem Subjekt. [DIE04]

5.2.6. (Rechts-)Verbindlichkeit (engl. non repudiation)

„Ein System gewährleistet die Verbindlichkeit bzw. Zuordenbarkeit einer Menge von Aktionen zu einem Subjekt, wenn es nicht möglich ist, dass ein Subjekt im Nachhinein die Durchführung einer solchen Aktion abstreiten kann.“ [ECK06]

Die Verbindlichkeitseigenschaft ist besonders in dem rasant wachsenden Bereich des elektronischen Handels (engl. E-Commerce) und der elektronischen Geschäfte (engl. E-Business) von erheblicher Bedeutung, um die Rechtsverbindlichkeit getätigter, geschäftlicher Transaktionen (Käufe, Verträge etc.) zu garantieren. [ECK06] Diese Anforderung lässt sich durch Message Authentification Codes oder den Einsatz digitaler Signaturen erfüllen. Die Verbindlichkeit ist aber auch allgemein bei der Nutzung von Systemressourcen in Mehrbenutzersystemen oder auch zunehmen bei Grid-, Peer-to-Peer und Cloud-Computing von Interesse. Beispiele relevanter Aktionen sind hier der Verbrauch von Rechenzeit oder die Nutzung teurer Ein/Ausgabe-Geräte. Mit der Verbindlichkeitseigenschaft ist die Forderung nach Zuordenbarkeit (engl. accountability) unmittelbar verbunden. Dies erfordert Maßnahmen zur Überwachung (engl. Audit) sowie zur Protokollierung einzelner Benutzeraktivitäten. [SCL06]

5.2.7. Anonymisierung und Pseudomisierung

Aus Gründen des Datenschutzes und der Wahrung der Privatsphäre von Personen haben sich die Schutzziele Anonymisierung und Pseudomisierung entwickelt. Durch mobile Endgeräte und die ständige Vernetzung dieser mit dem Internet haben sich neue Möglichkeiten und Wege für Datensammler eröffnet. Subjekte sollen daher nur dann identifizierbar sein wenn dies gerade erforderlich ist, ansonsten sollten sie anonym bleiben.

„Unter der Anonymisierung versteht man das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft, einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können. Eine schwächere Form der Anonymisierung stellt die Pseudomisierung dar.“ [BEA10] Dabei handelt es sich um das Verändern personenbezogener Daten durch eine Zuordnungsvorschrift (z.B. die Verwendung von Pseudonymen) derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse ohne Kenntnis oder Nutzung der Zuordnungsvorschrift nicht mehr einer natürlichen Person zugeordnet werden können. [ECK06]

Bei jedem Internetzugriff (z.B. beim Surfen) fallen eine Vielzahl von Daten über persönliche und sachliche Verhältnisse des Nutzers an, obwohl diese häufig für die Abwicklung der Dienstprotokolle nicht erforderlich sind. Beispiele dafür sind die IP-Adresse des Nutzers, seine Herkunftsadresse, die URL der zuvor geladenen Web-Seite oder auch Datum und Uhrzeit des Zugriffs. Diese Daten können sowohl von Angreifern (Dritten), die die Datenleitung abhören, als auch von Dienst-Anbietern dazu verwendet werden, unautorisiert Profile über Nut-

zer zu erstellen. Anonymisierungsdienste versuchen, durch Kombinationen aus Vermeidungs- (z.B. Daten unterdrücken) und Verschleierungstechniken (z.B. Ersetzen durch Standardmuster, Verschlüsseln) eine Anonymisierung oder Pseudomisierung zu erzielen. Pseudomität erzielt man durch die Einführung von Pseudonymen, sodass die Identität eines Subjekts zwar einem vertrauenswürdigen Dritten bekannt ist, nicht jedoch jedem Kommunikationspartner. Im engeren Sinn zielen Anonymisierungstechniken darauf ab, Aufenthaltsorte und Kommunikationsbeziehungen so zu verschleiern, dass keine Bewegungs-, Kommunikations- oder Zugriffsprofile einzelner Benutzer durch unautorisierte Dritte erstellt werden können. [ECK06]

5.3. Bedrohung, Angriff, Schaden

Aus Sicht der IT-Sicherheit ist es wünschenswert, dass prinzipiell jedes IT-System sicher ist. Doch wie konstruiert man sichere IT-Systeme? Prinzipiell ist zu sagen, dass angesichts der Vielzahl von IT-Systemen, Diensten und Kommunikationsnetzwerken sowie den komplexen Wechselwirkungen dieser Komponenten keine 100% sicheren IT-Systeme möglich sind. Es wird praktisch immer Schwachstellen oder (noch) unentdeckte Programmierfehler geben. Daher bleibt die Absicherung und Beseitigung von Schwachstellen eine Daueraufgabe der IT-Sicherheit. Dennoch kann man indem man einige Punkte bei der Konstruktion sicherer Systeme befolgt, die Gefahr von Schwachstellen minimieren. [POG07]

5.3.1. Sichere IT-Systeme

Grundsätzlich sollte man bei der Planung und Umsetzung sicherer IT-Systeme vier Phasen iterativ durchlaufen:

- Planungsphase, von der Problemstellung zum Maßnahmenkatalog
- Ausführungsphase, vom Maßnahmenkatalog zum funktionsfähigen System
- Prüfungsphase, Analyse des funktionsfähigen System ergibt Evaluierungskatalog
- Anpassungsphase, Evaluierungskatalog als Grundlage für neue Problemstellung

Die Planungsphase umfasst eine Bedrohungs- und Risikoanalyse sowie die Aufstellung eines Sicherheitsmodells. Bedrohungen und mögliche Schwachstellen eines Systems müssen identifiziert werden und im Zuge der Risikoanalyse werden die geschätzte Eintrittswahrscheinlichkeit und das zu erwartende Schadensausmaß hinzugefügt. Im entwickelten Sicherheitsmodell wird entschieden welche Sicherheitsanforderungen an das zu konstruierende IT-System zu stellen sind. Beispielsweise muss bei einer öffentlich zugänglichen Datenbank das Schutzziel Vertraulichkeit nicht berücksichtigt werden, das Schutzziel Integrität hingegen schon (alle Benutzer dürfen alles Lesen, aber die Informationen müssen korrekt sein). Hilfreich können hier Kriterienkataloge sein (z.B. Common Criteria oder BSI-Grundschutzhandbuch). Im Sicherheitsmodell sollten die folgenden Sicherheitsgrundfunktionen behandelt werden und deren Umsetzung individuell für jedes zu konstruierende sichere IT-System festgelegt werden:

Identifikation und Authentifizierung (Möglichkeit Subjekte und Objekte eindeutig identifizieren zu können, sowie Methode wie dies erfolgt, Schutzziele: Vertraulichkeit, Zuordenbarkeit), *Rechteverwaltung* sowie *Rechteprüfung* (Den Benutzern nur die Rechte geben, die sie wirklich benötigen und festlegen wann und wie diese Rechte überprüft werden sollen; Schutzziel: Vertraulichkeit), *Protokollierung* (Um Änderungen nachvollziehen zu können; Schutzziele: Zuordenbarkeit, Verbindlichkeit), *Wiederaufbereitung* (Bereinigung gemeinsam benutzter Betriebsmittel, Schutzziel: Vertraulichkeit), *Gewährleistung der Funktionalität* (Entscheidung welche Funktionalität mit welcher Priorität verfügbar sein muss, soft-failure Pinzip; Schutzziel: Verfügbarkeit). [POG07]

5.3.2. Terminologie Schadensmodell

Kapitel 5.3.1 beschreibt wichtige Grundschritte um ein Schadensmodell aufzustellen. Stellt man ein Schadensmodell auf wird man dazu zwangsläufig einige Begriffe benötigen, die im Folgenden kurz definiert werden:

Bedrohung (threat)

Mögliche unerwünschte oder unberechtigte Aktivitäten, die ein IT-System negativ beeinflussen können: Mögliche Verletzung der IT-Sicherheit und damit Ursache für ein unerwünschtes Ereignis [laut ISO/IEC 2382-14]. Eine Bedrohung zielt darauf ab Schwachstellen auszunutzen. Sicherheitsmaßnahmen sind materielle oder personelle Maßnahmen um eine Schwachstelle gegen eine Bedrohung abzusichern. Bedrohungen werden unterschieden nach: [POH04]

- Höhere Gewalt
- Fahrlässigkeit
- Absichtlichem Vorgehen (z.B. Spionage oder Sabotage)

Schwachstelle (vulnerability)

Sicherheitsrelevante Fehler eines IT-Systems – speziell eines Mechanismus [ISO/IEC 2003]. Schwachstellen können meist auf Fehler zurückgeführt werden. Fehler können sporadisch oder permanent wirksam sein. Sie können erkannt oder nicht erkannt sein. Synonym: Verwundbarkeit, Sicherheitslücke. Z.B. Diebstahlgefahr, Stromausfallgefahr oder Software Bug. [POH04]

Gefahr (danger)

Mögliches Eintreten einer Bedrohung gegen ein IT-System – unabhängig vom Vorhandensein eventueller Schwachstellen oder Sicherheitsmaßnahmen. [POH04]

Risiko (risk)

Quantitative Bewertung der Möglichkeit einer Ausnutzung einer Schwachstelle und damit Möglichkeit der Schadensverursachung [laut ISO/IEC 2002b und ISO/IEC 2003]. Oder auch Verhältnis aus Eintritt eines Schadensereignisses und dem bei Ereigniseintritt zu erwartenden Schadensausmaß [laut DIN/IEC 880] sowie Kombination aus Eintrittswahrscheinlichkeit und Konsequenzen eines Ereignisses [laut ISO/IEC 2002a]. [POH04]

Schaden (damage)

Minderung des Werts eines Objekts. Allgemein werden materielle (Sach- und Personenschäden) und immaterielle Schäden (z.B. Informationsdiebstahl) unterschieden. Materielle Schäden lassen sich auch mit dem Nicht-Erreichen von Sachzielen beschreiben: Dem Verlust der Verfügbarkeit, der Integrität, der Vertraulichkeit oder anderer Sachziele [laut BSI 2003]. Ein Schaden kann also sowohl im IT-System auftreten als auch außerhalb (Umwelt). [POH04]

Abbildung 20 veranschaulicht die zuvor genannten Begriffe anhand eines generischen Schadensmodells:

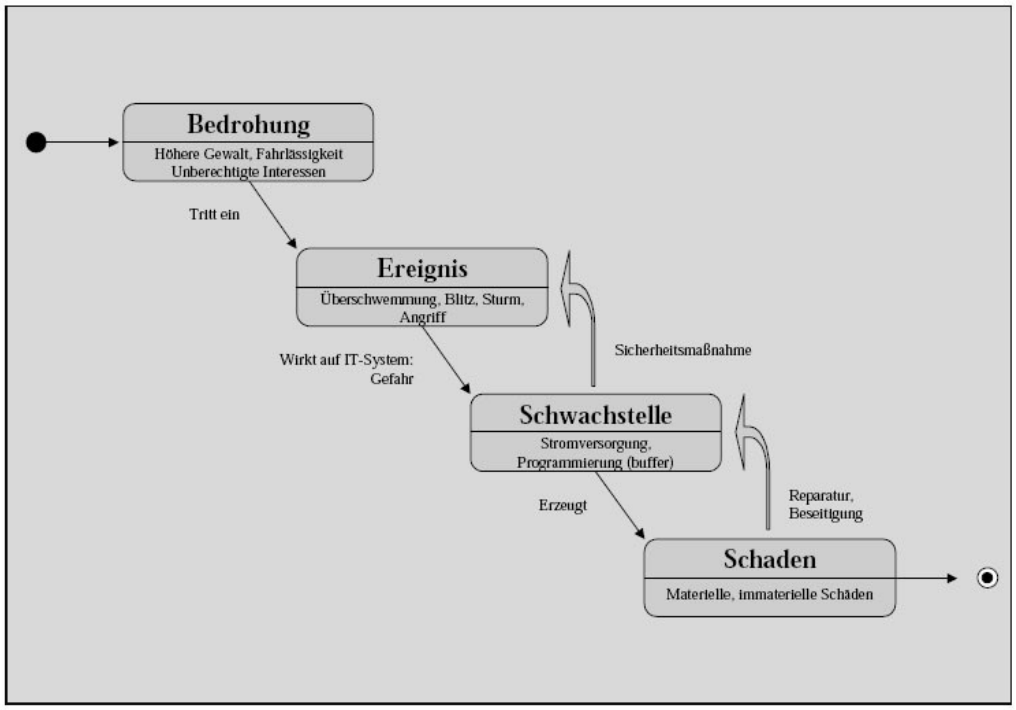

Abbildung 20: Generisches Schadensmodell

[POH04]

5.4. Generische Angriffsarten

Im Punkt „Sicherheitsüberlegungen“ jeweils bei den einzelnen E-Mail Protokollen, wird auf die für das jeweilige Protokoll spezifische, sicherheitstechnische Bewertungen eingegangen. Dieses Kapitel hingegen versucht die E-Mail Protokolle zueinander in Beziehung zu setzen und einen sicherheitstechnischen Vergleich aller Protokolle zu ermöglichen. Dazu wurden unter anderem einige pauschal anwendbare Angriffsarten identifiziert:

Sniffing

Unter (Packet) Sniffing versteht man das Mitlesen von Paketen im Netzwerk. Dies ist nur innerhalb des gleichen subnets möglich, indem sich auch das Sniffing Opfer befindet. Für diesen Zweck ist spezielle Software, so genannte Sniffer, verfügbar. Sniffer stellen einen eigenen Treiber für die Netzwerkkarte des Computers, den Capture Treiber, bereit. Dieser schaltet die Netzwerkkarte in den promiscuous mode, indem der Computer alle Netzwerkpakete mitliest, auch die, die eine andere MAC Adresse besitzen und daher nicht für diesen Computer bestimmt wären und normalerweise verworfen würden. Einfache Sniffer konnten nur die Daten von textuellen Protokollen darstellen. Binäre Protokolle müssen zuerst vom Sniffer aufbereitet werden um sie für Menschen lesbar zu machen. Daher bot früher die Verwendung von Binär Protokollen einen gewissen Schutz gegen Sniffing. Mittlerweile beherrschen die wichtigsten Sniffer aber zusätzlich auch die Fähigkeit Binär Protokolle lesbar aufzubereiten.

Der einzig verlässliche Schutz gegen Sniffing ist die Verschlüsselung der Paketdaten. Die verschlüsselten Daten können zwar vom Sniffer mitgelesen werden, sind aber ohne passenden Schlüssel für die Entschlüsselung wertlos. In der Sicherheitsmatrix wird bewertet ob das jeweilige Protokoll über eine integrierte Verschlüsselung verfügt.

Replay- und Janus-Angriff

Ein Replay-Angriff wird verwendet um Daten, die von einem Kommunikationsprotokoll zwischen den verschiedenen Teilnehmern übertragen werden, aufzuzeichnen und zu einem späteren Zeitpunkt erneut ins Netzwerk einzuspielen. Ein Janus-Angriff (auch Man-in-the-middle Angriff genannt) stellt quasi eine Erweiterung des klassischen Replay-Angriffs dar. Hierbei nimmt der Angreifer logisch oder physisch eine Stellung zwischen den Kommunikatonspartnern ein um die Pakete nicht nur aufzeichnen (sniffing) sondern auch direkt abfangen zu können - sie bei Bedarf zu manipulieren - und sie anschließend wieder in die Verbindung einzuspielen. Das wiedereingespielte Paket vermittelt dem Opfer den Eindruck, dass es von seinem normalen Kommunikationspartner kommt. Wurde beispielsweise ein Paket aufgezeichnet, welches für den Login-Vorgang verwendet wurde (enthält Benutzernamen sowie Hashwert des Passworts), so ist es dem Angreifer möglich die Identität dieses Kommunikationsteilneh-

mers zu übernehmen. Er kann damit unter dem Deckmantel einer fremden Identität diverse Aktionen ausführen und Schaden anrichten.

Sowohl Replay- als auch Janus-Angriffe sind nur schwer zu entdecken, da der Angreifer legitime Pakete verwendet (und diese modifiziert). Eine mögliche Gegenmaßnahme für Replay Angriffe ist die Verwendung von Nonces. Eine Nonce (i. d. Kryptographie, Abk. für „used only once“ oder „number used once“) ist ein einmalig erstellter Wert, der beispielsweise vor dem Login angefordert werden muss. Der User sendet diesen Wert mit seiner Login Nachricht an seinen Kommunikationspartner. Dieser akzeptiert die Nachricht nur, wenn sie die von ihm ausgesandte Nonce enthält, dies macht Replay-Angriffe zwecklos, schützt aber nicht vor Janus-Angriffen. Das TCP Protokoll erschwert Replay Angriffe, durch die Verwendung von Sequenznummern. Gemeinsam mit einem Message Authentification Code (MAC) stellt dies einen wirksamen Schutz gegen Replay-Angriffe dar. Einen vollwertigen Schutz gegen Janus-Angriffe und gegen das Mitlesen von Daten bieten aber nur Protokolle die Verschlüsselung und Authentifizierung unterstützen. Dies und ob TCP oder UDP wird in der Sicherheitsmatrix bewertet (Bezeichnung in der Matrix: Replay Angriff).

Denial-of-Service Angriff (DoS-Ang.)

Eine Denial-of-Service Attacke ist ein Nichtverfügbarkeitsangriff und zielt darauf ab eine Ressource (Protokoll, Website, Rechner, etc.) für Andere unverfügbar zu machen. Dies wird durch Überlastung (flooding) erreicht und führt dazu, dass eine Ressource keine Kapazitäten mehr hat andere Zugriffe, als die des Angreifers, anzunehmen.
Ein klassischer DoS Angriff ist ein SYN-Flood Angriff. Diese Angriffsart lässt sich bei allen Protokollen anwenden, die TCP als Übertragungsprotokoll verwenden. Der Angreifer sendet dabei TCP/SYN Pakete an das Opfer, das Opfer sendet TCP/SYN-ACK Pakete zurück und wartet auf die Antwort des Angreifers. Diese Antwort erfolgt allerdings nie. Die „halb“ geöffnete Verbindung bindet Ressourcen des Servers. Dies geschieht so lange bis der Server keine freien Ressourcen mehr hat und keine weiteren Verbindungen mehr annehmen kann.

Protokolle, die TCP als Übertragungsprotokoll verwenden sind daher deutlich anfälliger für Denial-of-Service Angriffe als Protokolle, die UDP als Übertragungsprotokoll verwenden.

5.5. Sicherheitsmatrix

In diesem Kapitel wird eine Sicherheitsmatrix vorgestellt. Diese kann einerseits als kompakte Übersichtstabelle der behandelten E-Mail Protokolle verstanden werden, andererseits sind in dieser Tabelle auch zusätzliche Informationen enthalten. Die folgende Tabelle listet spaltenweise die E-Mail Kernprotokolle auf. Zeilenweise sind jeweils Eckdaten der Protokolle angegeben und ob gewisse Kriterien, beispielsweise die Anfälligkeit für gewissen Angriffe, zutreffen. Die unterschiedlichen Angriffsarten wurden im vorherigen Kapitel 5.4 Generische Angriffsarten erklärt.

Protokoll **Auf Seite…**	**POP3** **28**	**IMAP4** **30**	**SMTP** **33**
Verbreitungsgrad	Sehr hoch	Sehr hoch	Sehr hoch
Schadenspotenzial	Sehr hoch	Sehr hoch	Sehr hoch
Verwendete TCP-Ports	110 und 995 (encr.)	143 und 993 (encr.)	25, 587, 465 (encr.)
Verwendete UDP-Ports	-	-	-
Einordnung Textprotokoll	Ja	Ja	Ja
Einordnung Binärprotokoll	Nein	Nein	Nein
Eingebaute Verschlüsselung	Opt	Opt	Nein
Eingebaute Authentifizierung	Opt	Ja	Opt
Anfällig für Sniffing?	Teilw	Teilw	Ja
Anfällig für Replay Angriff?	Teilw	Teilw	Ja
Anfällig für Denial-of-Service Angriff?	Ja	Ja	Ja

Legende:

Opt… Optional (Es existiert eine Protokollvariante die die Eigenschaft unterstützt)

Teilw… Teilweise (Das Protokoll ist je nach eingesetzter Protokollvariante für den Angriff anfällig oder nicht)

6. Fazit

Aktuell nimmt an vielen Orten der Welt die Bestrebung zu, die Meinungsfreiheit dahingehend einzuschränken, dass Kritiken am aktuellen politischen Regime möglichst unterdrückt werden, damit sie sich nicht in der Öffentlichkeit verbreiten können. Sei es in der arabischen Welt durch die Umbrüche des sogenannten „arabischen Frühlings“. Sei es in asiatischen Ländern wo autoritäre Regierungen versuchen negative Berichte zu unterdrücken, da diese schnell Unruhe in der Bevölkerung verbreiten könnten. Sei es in Eurasien wo einzelne Regierungen die Meinungsfreiheit einschränken wollen um sich für die nächste Wahl ins richtige Licht zu rücken. Sei es in Teilen Afrikas wo Regime alles dafür tun um zu verschleiern in welchen Töpfen der Staatshaushalt landet. Oder sei es in Amerika wo Bürger sich dem Überwachungsstaat entziehen wollen. Egal wo auf der Welt, Gründe anonym kommunizieren zu wollen gibt es viele. Durch die aktuelle Entwicklung gewinnt das Konzept der Anonymous Remailer zunehmend an Bedeutung.

Selbstverständlich sind Anonymous Remailer nicht dafür gedacht die allgemeine E-Mail Kommunikation ersetzen zu wollen. Sie sind vielmehr ein probates Hilfsmittel um einzelnen Personen oder Personengruppen - die Konsequenzen befürchten müssen, wenn Sie ihr Wissen teilen - dennoch das Kommunikationsmittel E-Mail zur Verfügung zu stellen. Es gibt viele Arten sein Wissen anderen Personen zugänglich zu machen. Die elektronische Kommunikation mittels E-Mail hat sich jedoch als effektives, schnelles und vor allem zielgerichtetes Mittel erwiesen Nachrichten zu verbreiten. In dieser Masterarbeit wurde eine freie Software entwickelt die eine Anonymous Remailer Funktionalität implementiert, genauer gesagt einen Cypherpunk Anonymous Remailer (Type I). Darüber hinaus enthält die Software jedoch auch Funktionen die weit über diejenigen eines gewöhnlichen Type I Anonymous Remailers hinausgehen. Im Unterschied zu einem Type II Remailer – bei dem ein spezielles Programm für das Absenden von E-Mails verwendet werden muss – nimmt der Type I Remailer E-Mail Nachrichten entgegen, die mit gewöhnlichen E-Mail Programmen gesendet wurden. Diese Flexibilität verringert normalerweise die Anonymität, da Angreifer den Anonymous Remailer von außen beobachten könnten. Dadurch könnten sie anhand der Nachrichtengröße und/oder des Eingangs- und Ausgangszeitpunkts eine Zuordnung, trotz der Anonymisierung, vornehmen.

Um dies zu verhindern implementiert der Anonymous Remailer eine Funktion um die Nachrichtengröße zu verändern. Es werden künstlich erzeugte Nachrichten-Header angehängt und an den Nachrichtentext wird zufällig generierter Text angehängt. Des Weiteren werden Attachements entfernt und durch die Anonymisierung werden Header und Body Elemente ggf. verändert. Diese Maßnahmen verändern die Größe der Nachrichten, was einen reinen Größenvergleich fehlschlagen lässt. Der Zeitvergleich zwischen Eintreffen der Nachricht beim Remailer und dem Absenden der anonymisierten Nachricht könnte ebenfalls eine Zuordnung

ermöglichen. Dieser wird durch zufälliges Verzögern der Nachrichten im Anonymous Remailer erheblich erschwert bzw. bei einer genügend hohen E-Mail Frequenz verunmöglicht.

In der Masterarbeit wurde ein besonderer Fokus auf das Feld des „Natural-Language-Processings“ gelegt. Dieses Forschungsfeld beschäftigt sich mit der Sprachverarbeitung, genauer gesagt, es sollen Texte semantisch analysiert werden können. Für die Masterarbeit ist ein Teilgebiet des „Natural-Language-Processings“ von Interesse, die „Named-Entity-Recognition“. Darunter versteht man das Erkennen und Extrahieren von benannten Objekten in einem Text. Solche Objekte können beispielsweise, Orte, Organisationen, Flüsse oder Personennamen sein. Im Gegensatz zu einem gewöhnlichen Type I Anonymous Remailer soll der Type I Anonymous Remailer der Masterarbeit zusätzlich in der Lage sein den eigentlichen Nachrichten-Text eines E-Mails zu filtern. Potenziell ‚verräterische‘ Elemente wie beispielsweise Personennamen, Orte und Organisationen werden optional vom Anonymous Remailer erkannt und on-the-fly anonymisiert.

Die Entwicklung des Anonymous Remailers (Type I) verlief nicht reibungslos. Es stellte sich die Frage mit welcher Programmiersprache dieser umgesetzt werden soll und welches Framework für die Umsetzung der komplexen Natural-Language-Processing Anforderungen verwendet werden soll? Aufgrund meiner beruflichen Nähe zu .NET Entwicklungen tendierte ich dazu C# als Programmiersprache zu verwenden. Nachdem ein erster Prototyp des Anonymous Remailers entwickelt war, konnte die Suche nach passenden Natural-Language-Processing Framework beginnen um diese anschließend im Anonymous Remailer Prototypen testen zu können. Es stellte sich - im Nachhinein betrachtet - die Entscheidung zuerst einen Prototyp zu programmieren und dann im Anschluss daran ein passendes NLP Framework zu finden, leider als Fehlentscheidung im Ablauf dar.

Das Finden eines effektiven und passenden Natural-Language-Processing Frameworks ist ein Kernbestandteil der Masterarbeit. Leider verlief die Suche nach frei verwendbaren und effektiven Natural Language Processing Frameworks sehr ernüchternd. Bei der Suche stieß ich zuerst auf das Proxem Antelope Framework (Link: https://www.proxem.com/en/technology/antelope-framework/). Es stellte sich aber heraus, dass dieses seit 2009 nicht mehr weiterentwickelt wurde und daher nicht mehr am aktuellen Stand der Technik ist. Ein weiteres Framework ist SharpNLP (Link: https://sharpnlp.codeplex.com/). Es ist eine C# Portierung des bekannten auf JAVA basierenden OpenNLP Frameworks. Es verwendet Maximum Entropy Modeling und enthält eine Name, Organisation, Ort, Zeit, Datum und Prozent Klassifizierung. Dieses Framework schien geeignet, jedoch erwies sich die Installation und Portierbarkeit als eher mangelhaft. Die Installation verlief in drei Teilen: Zuerst erfolgte die Installation der Maximum Entropy Models, bestehend aus 39 nbin Dateien (portierte OpenNLP Klassen), die die Grundlage für SharpNLP darstellen. Anschließend folgt das zu installierende Wörterbuch der Princeton

Universität „Wordnet“. Abschließend erfolgt die Installation der SharpNLP binaries, welche das eigentliche Programm enthalten. Die anschließenden Tests ergaben, dass vor allem der NameFinder – der für die Masterarbeit essentielle Teil - unzureichend arbeitet. Ein Beispieltext der unten im Screenshot abgebildet ist, ergab folgende Erkennungsleistung:

Gefundene Orte: Saffron Park, London, Austria
Nicht gefundene Orte: Linz
Gefunden Personen: Robert
Nicht gefunden Personen: Anne, Markus, Sarah

In der folgenden Abbildung wird ein exemplarischer Test mit SharpNLP gezeigt:

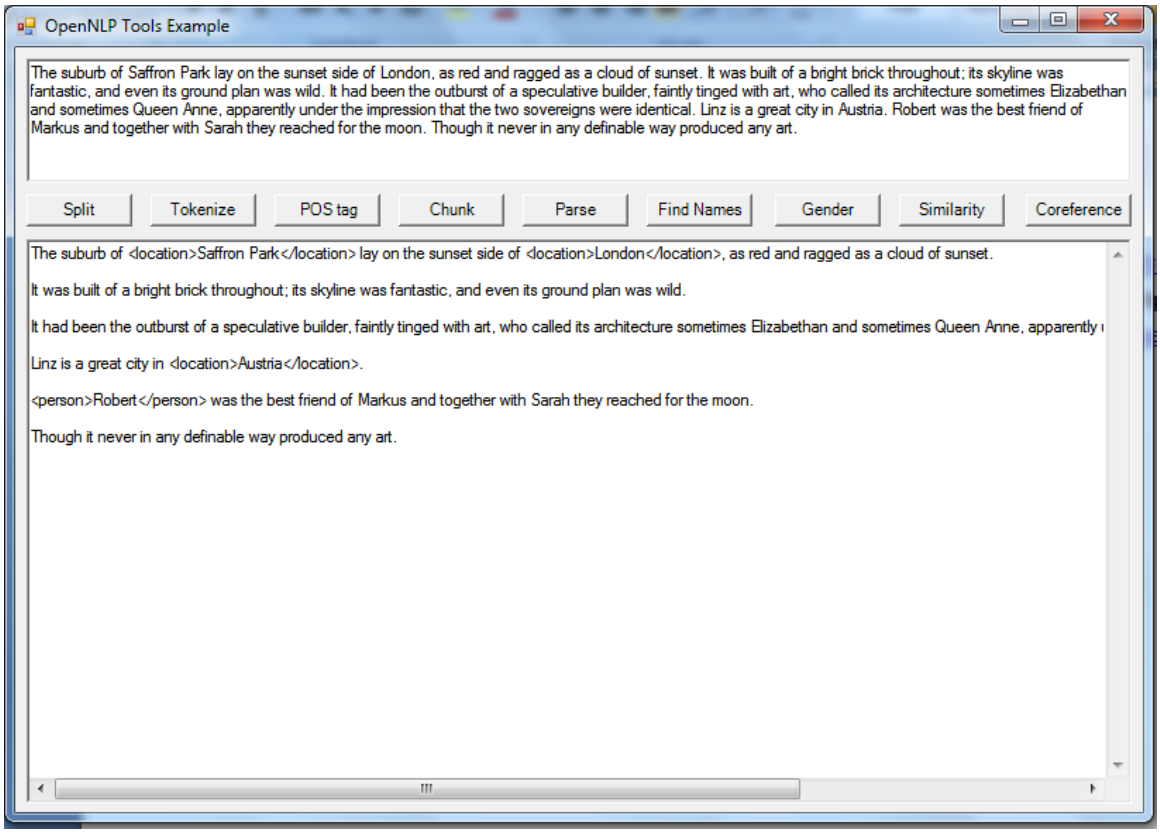

Abbildung 21: SharpNLP system test

Da auch bei anderen Beispieltexten die Erkennungsleistung unterdurchschnittlich ausfiel, blieb leider als einzige objektive Möglichkeit übrig, das SharpNLP Framework zu verwerfen. Weitere gefundene Frameworks für C# waren entweder noch in Entwicklung, nicht dokumentiert oder nicht mehr aktuelle Portierungen von namhaften Java Frameworks.

Dies hatte zur Konsequenz dass nun ein Prototyp mit Java entwickelt werden musste. Ziel war es ein verhältnismäßig schlankes Named-Entity-Recognition Java Framework zu finden, welches gleichzeitig hohe Erkennungsgenauigkeit bei vertretbaren Ressourceneinsatz liefert (es sollte auf einem günstigem Computer lauffähig sein und damit nicht sehr viel Arbeitsspeicher verbrauchen, was bei dieser Klasse von Algorithmen leider oft der Fall ist). Nach einer aus-

führlichen Suche bot sich der Stanford-Named-Entity-Recognizer (Link: http://nlp.stanford.edu/ner/) an. Das Stanford-NER ist ein eigenständiger Teil des Stanford Natural-Language-Processing Toolkits. Dadurch konnte der Anonymous Remailer verhältnismäßig „schlank“ gehalten werden, da keine für den Anonymous Remailer unnötigen Features verwendet werden mussten. Des Weiteren hat die Stanford Natural-Language-Processing Group einen hervorragenden wissenschaftlichen Ruf und forscht aktiv an der Verbesserung des Stanford-NER. Das Stanford-NER ist ein so genannter CRFClassifier. Die Software implementiert ein linear verkettetes „Conditional Random Field (CRF)“-Sequenzmodell. Durch diesen noch relativ neuen Forschungsansatz soll eine optimale Erkennungsleistung ermöglicht werden. Dies wurde in meinen Tests durchaus eindrucksvoll bestätigt. Bei mehreren Testbeispielen leistete sich das NER Framework keinen einzigen Fehler. Erst nach dem Einsatz von komplexen Testbeispielen wurden einige zu anonymisierende Elemente nicht klassifiziert. Durch entsprechendes „Training“ des Classifiers könnte man die Werte noch weiter verbessern. Dennoch ergibt sich auch ohne spezifisches Classifier-Training ein hervorragender Erkennungswert von 96% der zu klassifizierenden Items bei nur 0,8 % Falscherkennungen.

Used Classifier	Total no. of items	Correct items	False positive	Missing items	**% Classified**	**% False pos**
English	125	120	1	5	**96%**	**0,8%**

Im theoretischen Teil der Masterarbeit wird die grundsätzliche Sicherheit der verwendeten E-Mail Protokolle beurteilt und jeweils auf die Anonymisierungsmöglichkeiten dieser eingegangen. Dies ist insofern relevant da selbst der bestmöglich programmierte Anonymous Remailer angreifbar ist, wenn das darunter liegende System angreifbar ist. Daher werden die drei E-Mail Kernprotokolle, POP3, IMAP und SMTP erklärt und bewertet. Bei POP3 ist die secure Variante POP3S zu bevorzugen. Sie unterstützt das Einleiten der Verschlüsselung mit dem STARTTLS Befehl. Dabei muss der Client die Verschlüsselung unbedingt erzwingen, da sonst die Verschlüsselung – für ihn unbemerkt – deaktiviert werden kann, was ein Sicherheitsrisiko darstellt. Verschlüsselung ist insofern wichtig, da POP3 ein Klartext Protokoll ist, welches sowohl Benutzername als auch Passwort im Klartext sendet. Desweiteren ist der enthaltene Authentifizierungsmechanismus sehr einfach und es wird die Verwendung alternativer Authentifizierungsmethoden durch den AUTH Befehl - wie in RFC 1734 beschrieben - empfohlen.

IMAP unterstützt den STARTTLS Befehl für die Verschlüsselung. Mit einer verschlüsselten Verbindung ist auch der integrierte Authentifizierungsdienst sicher, da er Username und Passwort normalerweise im Klartext überträgt. Den Email Provider sollte man mit Bedacht wählen, da alle Mailboxen, inklusive deren Nachrichten, zentral auf dem Server des Anbieters

gespeichert werden. Der Anbieter sollte daher alle Datenschutz-Bestimmungen erfüllen. SMTP unterstützte in der ursprünglichen Version überhaupt keine Verschlüsselung. Erst mit Extended SMTP (ESMTP) wurde die Abfrage nach unterstützten Protokollerweiterungen nachgereicht. Mittlerweile bieten fast alle Server den STARTTLS Befehl an und damit ist es auch möglich SMTP verschlüsselt zu verwenden. Möchte der Client auch die E-Mail Nachrichten selbst verschlüsseln, so ist S/MIME für Nachrichtenverschlüsselung zu verwenden. Des Weiteren nehmen SMTP-Sever Nachrichten nicht von beliebigen Absendern entgegen, sondern sie akzeptieren nur diejenigen Nachrichten, bei denen sichergestellt ist, dass sie zum jeweiligen SMTP-Server passen. Dies geschieht z.B. anhand von IP/Domain-Kennungen und Benutzeranmeldedaten. Diese Praxis erschwert die anonyme Verwendung von SMTP erheblich. Daher wurden Anonymous Remailer entwickelt. Sie ermöglichen die anonyme Verwendung des Kommunikationsmediums E-Mail.

Mit dem Kapitel IT-Sicherheit soll verdeutlicht werden, dass die bei der Konstruktion sicherer IT-Systeme möglichen Ziele in einem Spannungsverhältnis zueinander stehen. Je nach Einsatzweck des Systems werden nicht alle Ziele gleich gewichten oder es müssen einzelne Ziele ausgenommen werden. Diese Ziele sind:

- **Vertraulichkeit (engl. confidentiality) (siehe Kapitel 5.2.2)**
  „Ein System gewährleistet die Informationsvertraulichkeit, wenn es keine unautorisierte Informationsgewinnung ermöglicht.“ [ECK06]
- **Integrität (engl. integrity) (siehe Kapitel 5.2.3)**
  „Ein System gewährleistet die Datenintegrität, wenn es Subjekten nicht möglich ist, die zu schützenden Daten unautorisiert und unbemerkt zu manipulieren.“ [ECK06]
- **Verfügbarkeit (engl. availability) (siehe Kapitel 5.2.4)**
  „Ein System gewährleistet Verfügbarkeit, wenn authentifizierte und autorisierte Subjekte in der Wahrnehmung ihrer Berechtigungen nicht unautorisiert beeinträchtigt werden können.“ [ECK06]
- **Zuordenbarkeit (engl. accountability) (siehe Kapitel 5.2.5)**
  „Von jeder in einem IT-System ausgeführten Aktion (Vorgang, Prozess) muss während ihres Ablaufs und danach feststellbar sein, wem, d.h. welcher Instanz – insbesondere welcher Person – diese Aktion zuzuordnen ist. Es muss klar feststellbar sein, welches Subjekt sie ausgelöst und wer sie letztlich zu verantworten hat.“ [DIE04] Diese Forderung gilt nicht nur für die Aktionen selbst, sondern erst recht für deren Ergebnisse oder Auswirkungen.
- **(Rechts-)Verbindlichkeit (engl. non repudiation) (siehe Kapitel 5.2.6)**
  „Ein System gewährleistet die Verbindlichkeit bzw. Zuordenbarkeit einer Menge von Aktionen zu einem Subjekt, wenn es nicht möglich ist, dass ein Subjekt im Nachhinein die Durchführung einer solchen Aktion abstreiten kann.“ [ECK06]

- **Anonymisierung und Pseudomisierung (siehe Kapitel 5.2.7)**
  Subjekte sollen nur dann identifizierbar sein, wenn dies gerade erforderlich ist. Ansonsten sollten sie anonym bleiben. Anonymisierung ist die gänzliche Unkenntbarmachung der personenbezogenen Daten. Bei der Pseudomisierung werden personenbezogene Daten gegen Pseudonyme ersetzt. Diese enthalten keine personenbezogenen Daten mehr, können jedoch vom Subjekt, welches die Pseudomisierung durchgeführt hat, wieder in personenbezogene Daten umgewandelt werden.

Legt man beispielsweise Wert auf das Schutzziel „Anonymisierung und Pseudomisierung“, so muss man mit hoher Wahrscheinlichkeit auf die Schutzziele Zuordenbarkeit und (Rechts-)Verbindlichkeit verzichten, da sich ein anonymer Zugang und eine eindeutige Zuordnung zu einer Person gegenseitig ausschließen. Dieses Beispiel soll verdeutlichen, dass bereits bei der Planung eines IT-Systems der Einsatzweck des IT-Systems definiert wird und darauf basierend die Gewichtung der Schutzziele durchgeführt wird. Anschließend wird geplant womit sich die Schutzziele in der Praxis umsetzen lassen. Beispielsweise mit welcher Technik oder welchem Algorithmus. Das resultierende IT-System soll bei höchstmöglicher Sicherheit die geforderten Schutzziele bestmöglich abdecken.

Mit Hilfe von, mittels S/MIME oder ähnlicher Techniken, eingeführten Digitalen Signaturen und Verschlüsselungstechniken ist es möglich Vertraulichkeit, Integrität, Zuordenbarkeit und (Rechts-)Verbindlichkeit bei der E-Mail Kommunikation zu gewährleisten. Das Schutzziel der „Anonymisierung und Pseudomisierung“ konnte jedoch nicht erfüllt werden.

Erst Anonymous Remailer machten und machen es möglich auch beim Kommunikationsmedium E-Mail Anonymität zu gewährleisten.

Appendix A (Classified Text)

Dieses Kapitel enthält alle Textsamples, die für die Klassifizierungstests verwendet wurden und die jeweiligen Ergebnisse dieser Tests. Es werden dabei alle korrekt klassifizierenden Elemente gezählt (correct item = r). Ebenso werden nicht-erfolgte Klassifizierungen (missing item = m) und zuviel-erfolgte Klassifizierungen (false positive item = f) gezählt. Dabei sind zuerst diejenigen Texte gelistet die für den Test des deutschsprachigen Classifiers verwendet wurden und anschließend die Texte für den Test des englischsprachigen Classifiers.

Deutsch

1)
Lieber Peter,

Unter diesen Dokumenten ist auch die komplette Liste der von der NATO geplanten und durchgeführten "gezielten Tötungen". Im Militärjargon ist es die "Joint Prioritized Effect List" (JPEL). Diese Liste belegt, dass die Strategie der gezielten Tötungen nicht bloß als letztmögliches Mittel zur Bekämpfung der Taliban eingesetzt wurde, sondern Teil des regulären Vorgehens in diesem Krieg war.

Der "Spiegel" konnte Dokumente aus den Jahren 2009 bis 2011 einsehen, zu dieser Zeit war Stanley McChrystal ISAF-Kommandeur, der im Juni 2010 von David Petraeus abgelöst wurde. Schon 2009 schickte US-Präsident Barack Obama 33.000 zusätzliche Soldaten nach Afghanistan. Eine der blutigsten Phasen des Krieges begann. 2009 starben 2.412 Zivilisten, für ein Viertel dieser Toten waren NATO-Truppen und afghanische Sicherheitskräfte verantwortlich. Auch die Zahl der Einsätze gegen die Taliban nahm zu.

Hinter der Eskalation stand die Strategie von Petraeus: Er wollte die Gegner zuerst mittels einer "Säuberungsphase" schwächen, um anschließend das entstandene Machtvakuum mit lokalen Kräften zu füllen. Danach sollte eine Phase der Stabilisierung folgen.

Mfg
Robert Söllner
Quelle: http://derstandard.at/2000009887439/NATO-Einsatz-in-Afghanistan-Gezielte-Toetungen-waren-Alltag

Personen (r/f/m): 5/0/1 , Organisationen (r/f/m): 2/2/2 , Orte (r/f/m): 2/0/0

Lieber <I-PER>Peter</I-PER>,

Unter diesen Dokumenten ist auch die komplette Liste der von der <I-ORG>NATO</I-ORG> geplanten und durchgeführten "gezielten Tötungen". Im Militärjargon ist es die <I-ORG>"Joint Prioritized Effect List</I-ORG>" (<I-ORG>JPEL</I-ORG>). Diese Liste belegt, dass die Strategie der gezielten Tötungen nicht bloß als letztmögliches Mittel zur Bekämpfung der Taliban eingesetzt wurde, sondern Teil des regulären Vorgehens in diesem Krieg war.

Der "Spiegel" konnte Dokumente aus den Jahren 2009 bis 2011 einsehen, zu dieser Zeit war <I-PER>Stanley McChrystal</I-PER> ISAF-Kommandeur, der im Juni 2010 von <I-PER>David Petraeus</I-PER> abgelöst wurde. Schon 2009 schickte US-Präsident Barack Obama 33.000 zusätzliche Soldaten nach <I-LOC>Afghanistan</I-LOC>. Eine der blutigsten Phasen des Krieges begann. 2009 starben 2.412 Zivilisten, für ein Viertel dieser Toten waren <I-MISC>NATO-Truppen</I-MISC> und <I-MISC>afghanische</I-MISC> Sicherheitskräfte verantwortlich. Auch die Zahl der Einsätze gegen die Taliban nahm zu.

Hinter der Eskalation stand die Strategie von <I-PER>Petraeus</I-PER>: Er wollte die Gegner zuerst mittels einer "Säuberungsphase" schwächen, um anschließend das entstandene Machtvakuum mit lokalen Kräften zu füllen. Danach sollte eine Phase der Stabilisierung folgen.

Mfg
<I-PER>Robert Söllner</I-PER>
Quelle: http://derstandard.at/2000009887439/NATO-Einsatz-in-Afghanistan-Gezielte-Toetungen-waren-Alltag

2)
Der berüchtigte Abmahn-Anwalt Thomas Urmann will seine Zulassung als Anwalt freiwillig zurückgelegt haben. Das sagt Urmann in einem Gespräch mit der deutschen Zeitung "Die Welt". Urmann tritt damit Berichten entgegen, die Anwaltskammer Nuernberg habe ihm seine Konzession wegen sittenwidrigen Verhaltens entzogen.

Als Grund für den Ruecktritt gibt Urmann an, selbst in einem Strafverfahren angeklagt zu sein. "Dann kann man meiner Meinung nach nicht mehr als Organ der Rechtspflege auftreten", so Urmann. Dem Juristen wird unter anderem wegen Insolvenzverschleppung der Prozess gemacht.

Urmann hatte im vergangenen Jahr für Aufsehen gesorgt, als er zigtausende Nutzer der Pornoplattform RedTube abmahnte. Sie sollten für Urheberrechtsverletzungen bezahlen. Laut Ansicht vieler IT-Rechtsexperten gebe es keine gesetzliche Grundlage für solche Abmahnaktionen, manche bezeichnen sie sogar als sittenwidrig. Bereits zuvor hatte Urmann die Netzgemeinde veraergert, da er viele Betreiber kleiner Internetshops wegen Fehlern in deren Impressum abmahnen wollte.

Quelle: http://derstandard.at/2000009927970/RedTube-Mahnanwalt-will-Zulassung-freiwillig-zurueckgelegt-haben

Personen (r/f/m): 4/1/3 , Organisationen (r/f/m): 0/0/1 , Locations (r/f/m): 1/0/0

Der berüchtigte Abmahn-Anwalt <I-PER>Thomas Urmann</I-PER> will seine Zulassung als Anwalt freiwillig zurückgelegt haben. Das sagt <I-PER>Urmann</I-PER> in einem Gespräch mit der <I-MISC>deutschen</I-MISC> Zeitung "Die Welt". Urmann tritt damit Berichten entgegen, die Anwaltskammer <I-PER>Nuernberg</I-PER> habe ihm seine Konzession wegen sittenwidrigen Verhaltens entzogen.

Als Grund für den Ruecktritt gibt Urmann an, selbst in einem Strafverfahren angeklagt zu sein. "Dann kann man meiner Meinung nach nicht mehr als Organ der Rechtspflege auftreten", so <I-PER>Urmann</I-PER>. Dem Juristen wird unter anderem wegen Insolvenzverschleppung der Prozess gemacht.

<I-PER>Urmann</I-PER> hatte im vergangenen Jahr für Aufsehen gesorgt, als er zigtausende Nutzer der Pornoplattform RedTube abmahnte. Sie sollten für Urheberrechtsverletzungen bezahlen. Laut Ansicht vieler <I-MISC>IT-Rechtsexperten</I-MISC> gebe es keine gesetzliche Grundlage für solche Abmahnaktionen, manche bezeichnen sie sogar als sittenwidrig. Bereits zuvor hatte Urmann die Netzgemeinde verärgert, da er viele Betreiber kleiner Internetshops wegen Fehlern in deren Impressum abmahnen wollte.

Quelle: http://derstandard.at/2000009927970/RedTube-Mahnanwalt-will-Zulassung-freiwillig-zurueckgelegt-haben

---

3)
Der Arbeitskampf der deutschen Dienstleistungsgewerkschaft Verdi beim Online-Versandhändler Amazon hat für den Verdi-Chef grundsätzliche Ausstrahlung weit über den deutschen Ableger des US-Giganten hinaus. "Es ist ein fundamentaler Konflikt", sagte Frank Bsirske der Deutschen Presse-Agentur in Berlin.

"Es geht darum, wie im digitalen Zeitalter die Zukunft der Arbeitsbeziehungen gestaltet werden soll", sagte Bsirske. "Hier sollen amerikanisierte Arbeitsbeziehungen mit einer Ablehnung von Gewerkschaften als Verhandlungspartner nach Europa exportiert werden."

Kurzfristig zeichne sich auch nach den jüngsten Streiks bei Amazon kein Durchbruch ab. Verdi habe zuletzt sechs Standorte in der Bundesrepublik einbezogen, Arbeitsniederlegungen habe es auch in Frankreich gegeben

Quelle: http://derstandard.at/2000009928474/Gewerkschaft-Amazon-Arbeitskampf-hat-grundsaetzliche-Ausstrahlung

Personen (r/f/m): 4/0/1 , Organisationen (r/f/m): 0/0/0 , Locations (r/f/m): 7/0/1

Der Arbeitskampf der <I-MISC>deutschen</I-MISC> Dienstleistungsgewerkschaft <I-PER>Verdi</I-PER> beim Online-Versandhändler Amazon hat für den Verdi-Chef grundsätzliche Ausstrahlung weit über den <I-MISC>deutschen</I-MISC> Ableger des <I-MISC>US-Giganten</I-MISC> hinaus. "Es ist ein fundamentaler Konflikt", sagte <I-PER>Frank Bsirske</I-PER> der <I-MISC>Deutschen</I-MISC> Presse-Agentur in <I-LOC>Berlin</I-LOC>.

"Es geht darum, wie im digitalen Zeitalter die Zukunft der Arbeitsbeziehungen gestaltet werden soll", sagte <I-PER>Bsirske</I-PER>. "Hier sollen amerikanisierte Arbeitsbeziehungen mit einer Ablehnung von Gewerkschaften als Verhandlungspartner nach <I-LOC>Europa</I-LOC> exportiert werden."

Kurzfristig zeichne sich auch nach den jüngsten Streiks bei Amazon kein Durchbruch ab. <I-PER>Verdi</I-PER> habe zuletzt sechs Standorte in der <I-LOC>Bundesrepublik</I-LOC> einbezogen, Arbeitsniederlegungen habe es auch in <I-LOC>Frankreich</I-LOC> gegeben

Quelle: http://derstandard.at/2000009928474/Gewerkschaft-Amazon-Arbeitskampf-hat-grundsaetzliche-Ausstrahlung

---

4)
Spitzenreiter im Vertrauensranking der Österreicher ist traditionell Bundespräsident Heinz Fischer, dessen Werte zwischen 45 Punkten und zuletzt im Dezember 52 Punkten variierten. Auf Platz zwei rangierte bis zu ihrem Ableben im Sommer Nationalratspräsidentin Barbara Prammer, SPÖ (plus 29 bzw. 31 Punkte). Ihre Nachfolgerin Doris Bures (SPÖ) kam da mit plus einem Prozent im Dezember nicht einmal in die Nähe, hatte aber gegenüber ihrer Werte als Infrastrukturministerin den Weg aus dem Minus (Tiefstwert 2014: minus elf) vollbracht.

Das vertrauenswürdigste Regierungsmitglied war das ganze Jahr über, zuletzt mit 22 Pluspunkten, Außenminister Sebastian Kurz (ÖVP). Deutlich verloren hat im Jahresverlauf Verteidigungsminister Gerald Klug (SPÖ), der ein Bundesheer-Sparpaket zu kommunizieren und verhandeln hatte: Von plus neun Punkten fiel er auf minus acht zurück.

Quellle: http://derstandard.at/2000009944628/Faymann-2014-im-Minus-Bonus-fuer-Mitterlehner

Personen (r/f/m): 5/0/0 , Organisationen (r/f/m): 4/0/0 , Locations (r/f/m): 0/0/1

Spitzenreiter im Vertrauensranking der Oesterreicher ist traditionell Bundespraesident <I-PER>Heinz Fischer</I-PER>, dessen Werte zwischen 45 Punkten und zuletzt im Dezember 52 Punkten variierten. Auf Platz zwei rangierte bis zu ihrem Ableben im Sommer Nationalratspraesidentin <I-PER>Barbara Prammer</I-PER>, <I-ORG>SPÖ</I-ORG> (plus 29 bzw. 31 Punkte). Ihre Nachfolgerin <I-PER>Doris Bures</I-PER> (<I-ORG>SPÖ</I-ORG>) kam da mit plus einem Prozent im Dezember nicht einmal in die Nähe, hatte aber gegenüber ihrer Werte als Infrastrukturministerin den Weg aus dem Minus (Tiefstwert 2014: minus elf) vollbracht.

Das vertrauenswürdigste Regierungsmitglied war das ganze Jahr über, zuletzt mit 22 Pluspunkten, Außenminister <I-PER>Sebastian Kurz</I-PER> (<I-ORG>ÖVP</I-ORG>). Deutlich verloren hat im Jahresverlauf Verteidigungsminister <I-PER>Gerald Klug</I-PER> (<I-ORG>SPÖ</I-ORG>), der ein Bundesheer-Sparpaket zu kommunizieren und verhandeln hatte: Von plus neun Punkten fiel er auf minus acht zurück.

Quellle: http://derstandard.at/2000009944628/Faymann-2014-im-Minus-Bonus-fuer-Mitterlehner

---

5)
Der Skiunfall hatte sich am Vortag im Ötschergebiet ereignet. Hundstorfer verkantete und kam ohne Fremdverschulden zu Fall, berichtete die Sprecherin des Ministers auf APA-Anfrage. Die Erstversorgung oblag dem Krankenhaus Scheibbs, dessen Personal der Ressortchef einen Dank ausrichtete. Behandelt wird Hundstorfer nunmehr im Wiener AKH.

Die österreichische Innenpolitik hätte den Sozialminister gerne rasch wieder auf den Beinen. Genesungswünsche richteten Kanzler Werner Faymann (SPÖ), Vizekanzler Reinhold Mitterlehner (ÖVP), FP-Generalsekretär Herbert Kickl und Grünen-Bundessprecherin Eva Glawischnig aus. Wissenschaftsminister Mitterlehner versicherte, dass der Patient im AKH in "bester medizinische Versorgung" sei.

Quelle: http://www.kleinezeitung.at/s/politik/innenpolitik/4628803/Beckenbruch_Hundstorfer-nach-Skiunfall-schwer-verletzt

Personen (r/f/m): 4/0/3 , Organisationen (r/f/m): 6/0/1 , Locations (r/f/m): 3/0/1

Der Skiunfall hatte sich am Vortag im Oetschergebiet ereignet. Hundstorfer verkantete und kam ohne Fremdverschulden zu Fall, berichtete die Sprecherin des Ministers auf <I-MISC>APA-Anfrage</I-MISC>. Die Erstversorgung oblag dem Krankenhaus <I-LOC>Scheibbs</I-LOC>, dessen Personal der Ressortchef einen Dank ausrichtete. Behandelt wird Hundstorfer nunmehr im <I-MISC>Wiener</I-MISC> <I-ORG>AKH</I-ORG>.

Die <I-MISC>Österreichische</I-MISC> Innenpolitik hatte den Sozialminister gerne rasch wieder auf den Beinen. Genesungswuensche richteten Kanzler <I-PER>Werner Faymann</I-PER> (<I-ORG>SPÖ</I-ORG>), Vizekanzler <I-PER>Reinhold Mitterlehner</I-PER> (<I-ORG>ÖVP</I-ORG>), <I-MISC>FP-Generalsekretär</I-MISC> <I-PER>Herbert Kickl</I-PER> und Grünen-Bundessprecherin <I-PER>Eva Glawischnig</I-PER> aus. Wissenschaftsminister Mitterlehner versicherte, dass der Patient im <I-ORG>AKH</I-ORG> in "bester medizinische Versorgung" sei.
Quelle: http://www.kleinezeitung.at/s/politik/innenpolitik/4628803/Beckenbruch_Hundstorfer-nach-Skiunfall-schwer-verletzt

English:

1)
As the 114th Congress prepares to convene on Tuesday, some senior members, who will not be returning, reflected on what they saw as necessary to the success of both lawmakers and the institution.

In interviews, Senator Tom Harkin, Democrat of Iowa; Senator Saxby Chambliss, Republican of Georgia; Representative Henry A. Waxman, Democrat of California; and Representative Jack Kingston, Republican of Georgia, pointed out the problems of recent years — failures that contributed to a steep decline in productivity on Capitol Hill as the gulf between the parties widened.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

Person (r/f/m): 4/0/0 , Organization (r/f/m): 1/0/0 , Location (r/f/m): 5/0/0

As the 114th <ORGANIZATION>Congress</ORGANIZATION> prepares to convene on Tuesday, some senior members, who will not be returning, reflected on what they saw as necessary to the success of both lawmakers and the institution.

In interviews, Senator <PERSON>Tom Harkin</PERSON>, Democrat of <LOCATION>Iowa</LOCATION>; Senator <PERSON>Saxby Chambliss</PERSON>, Republican of <LOCATION>Georgia</LOCATION>; Representative <PERSON>Henry A. Waxman</PERSON>, Democrat of <LOCATION>California</LOCATION>; and Representative <PERSON>Jack Kingston</PERSON>, Republican of <LOCATION>Georgia</LOCATION>, pointed out the problems of recent years failures that contributed to a steep decline in productivity on <LOCATION>Capitol Hill</LOCATION> as the gulf between the parties widened.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

---

2)
Longer still were the 15 years from his first hearing with tobacco company executives to the signing of a major tobacco control bill, giving the Food and Drug Administration the authority to regulate tobacco products in 2009.

“I always felt that you have to reach out to Republicans,” Mr. Waxman said. “When I was chairman of the health and environment subcommittee, I welcomed their comments. I wanted to hear what they had to say. Many times they would make criticisms or proposals that improved the legislation, and I welcomed that.”

Mr. Waxman collaborated with Senator Orrin G. Hatch, Republican of Utah, on a law that gave consumers access to inexpensive generic versions of brand-name drugs. The 1984 law “is a great example of two very different members of Congress working together to address a critical need,” Mr. Hatch said.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

Person (r/f/m): 4/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 1/0/0

Longer still were the 15 years from his first hearing with tobacco company executives to the signing of a major tobacco control bill, giving the <ORGANIZATION>Food and Drug Administration</ORGANIZATION> the authority to regulate tobacco products in 2009.

I always felt that you have to reach out to Republicans, Mr. <PERSON>Waxman</PERSON> said. When I was chairman of the health and environment subcommittee, I welcomed their comments. I wanted to hear what they had to say. Many times they would make criticisms or proposals that improved the legislation, and I welcomed that.

Mr. <PERSON>Waxman</PERSON> collaborated with Senator <PERSON>Orrin G. Hatch</PERSON>, Republican of <LOCATION>Utah</LOCATION>, on a law that gave consumers access to inexpensive generic versions of brand-name drugs. The 1984 law is a great example of two very different members of <ORGANIZATION>Congress</ORGANIZATION> working together to address a critical need, Mr. <PERSON>Hatch</PERSON> said.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

---

3)
There are fewer departing lawmakers than in the two previous Congresses, but many of those leaving are seasoned legislators. Among those exiting are Representative John D. Dingell, the Michigan Democrat and longest serving House member in history, and Senator Carl Levin, Democrat of Michigan and chairman of the Armed Services Committee, who spent 36 years in the Senate before casting his final vote on Dec. 16. They also served in Washington for 7 year together with Senator John Brigthing.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

Person (r/f/m): 3/0/0 , Organization (r/f/m): 3/0/0 , Location (r/f/m): 3/0/0

There are fewer departing lawmakers than in the two previous Congresses, but many of those leaving are seasoned legislators. Among those exiting are Representative <PERSON>John D. Dingell</PERSON>, the <LOCATION>Michigan</LOCATION> Democrat and longest serving <ORGANIZATION>House</ORGANIZATION> member in history, and Senator <PERSON>Carl Levin</PERSON>, Democrat of <LOCATION>Michigan</LOCATION> and chairman of the <ORGANIZATION>Armed Services Committee</ORGANIZATION>, who spent 36 years in the <ORGANIZATION>Senate</ORGANIZATION> before casting his final vote on Dec. 16. They also served in <LOCATION>Washington</LOCATION> for 7 year together with Senator <PERSON>John Brigthing</PERSON>.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

---

4)
After a victory in November on a Washington State ballot measure that will require broader background checks on gun buyers, groups that promote gun regulations have turned away from Washington and the political races that have been largely futile. Instead, they are turning their attention - and their growing wallets - to other states that allow ballot measures.

An initiative seeking stricter background checks for certain purchasers has already qualified for the 2016 ballot in Nevada, where such a law was passed last year by the Legislature then vetoed by the governor. Advocates of gun safety - the term many now use instead of “gun control” - are seeking lines on ballots in Arizona, Maine and Oregon as well.

I can’t recall ballot initiatives focused on gun policy, said Daniel Webster, the director of the Johns Hopkins Center for Gun Policy and Research.

Quelle: http://www.nytimes.com/2015/01/03/us/gun-control-groups-blocked-in-washington-turn-attention-to-states.html

Person (r/f/m): 1/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 6/0/0

After a victory in November on a <LOCATION>Washington State</LOCATION> ballot measure that will require broader background checks on gun buyers, groups that promote gun regulations have turned away from <LOCATION>Washington</LOCATION> and the political races that have been largely futile. Instead, they are turning their attention - and their growing wallets - to other states that allow ballot measures.

An initiative seeking stricter background checks for certain purchasers has already qualified for the 2016 ballot in <LOCATION>Nevada</LOCATION>, where such a law was passed last year by the <ORGANIZATION>Legislature</ORGANIZATION> then vetoed by the governor. Advocates of gun safety - the term many now use instead of gun control - are seeking lines on ballots in <LOCATION>Arizona</LOCATION>, <LOCATION>Maine</LOCATION> and <LOCATION>Oregon</LOCATION> as well.

I cant recall ballot initiatives focused on gun policy, said <PERSON>Daniel Webster</PERSON>, the director of the <ORGANIZATION>Johns Hopkins Center</ORGANIZATION> for Gun Policy and Research.

Quelle: http://www.nytimes.com/2015/01/03/us/gun-control-groups-blocked-in-washington-turn-attention-to-states.html

---

5)
If lawmakers are to break out of the partisan cycle, Mr. Kingston said, they need to avoid being inundated by their constituents in an increasingly digital world where members of Congress find themselves under immediate pressure as events unfold.

“If new members allow their base to control their behavior up here they are going to be miserable,” said Mr. Kingston, who has seen the rising influence of Tea Party activists in Houston on Republican lawmakers. “While the voters might be yelling and screaming at you to do something, that’s not your job.

“You have to look at all the information and then make the best determination as to what’s going to be best for America,” he said. “Sometimes you have to have disagreements with your own party along the way, and that is O.K.”

A similar sentiment was expressed by Mr. Harkin, who was the principal sponsor of the Americans With Disabilities Act of 1990, and 18 years later was the chief sponsor of a law that expanded disability rights again by overturning several FBI decisions.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

Person (r/f/m): 3/0/0 , Organization (r/f/m): 4/0/0 , Location (r/f/m): 2/0/0

If lawmakers are to break out of the partisan cycle, Mr. <PERSON>Kingston</PERSON> said, they need to avoid being inundated by their constituents in an increasingly digital world where members of <ORGANIZATION>Congress</ORGANIZATION> find themselves under immediate pressure as events unfold.

If new members allow their base to control their behavior up here they are going to be miserable, said Mr. <PERSON>Kingston</PERSON>, who has seen the rising influence of <ORGANIZATION>Tea Party</ORGANIZATION> activists in <LOCATION>Houston</LOCATION> on <ORGANIZATION>Republican</ORGANIZATION> lawmakers. While the voters might be yelling and screaming at you to do something, thats not your job.

You have to look at all the information and then make the best determination as to whats going to be best for <LOCATION>America</LOCATION>, he said. Sometimes you have to have disagreements with your own party along the way, and that is O.K.

A similar sentiment was expressed by Mr. <PERSON>Harkin</PERSON>, who was the principal sponsor of the Americans With Disabilities Act of 1990, and 18 years later was the chief sponsor of a law that expanded disability rights again by overturning several <ORGANIZATION>FBI</ORGANIZATION> decisions.

Quelle: http://www.nytimes.com/2015/01/03/us/politics/departing-lawmakers-lament-capitols-partisanship.html

---

6)
Dear Mrs. Welsch,
I want to thank you for the great weekend in Tirol. The snow was marvelous and our host, the family Steinbacher were very polite and hospitable. Even the Siemens people did not try to invite us to their homebase in Vienna, Austria.

I hope you also enjoyed our stay the same way we did. We already told the McCallesters how nice our holiday was and the said they are planning to visit Kitzbuehel too. They are great skiers and they will love it when they get the chance to freeride in the mountains with you.

My wife Angela wants to add that Dani should buy the discussed shoes in Salzburg and send it via packet to Auckland. Many thanks in advance for that. I’ll soon write you again when Angela and me arrive in Moscow.

Kind Greetings,
Patrick Rinch

Person (r/f/m): 6/0/1 , Organization (r/f/m): 1/0/0 , Location (r/f/m): 7/0/0

Dear Mrs. <PERSON>Welsch</PERSON>,
I want to thank you for the great weekend in <LOCATION>Tirol</LOCATION>. The snow was marvelous and our host, the family <PERSON>Steinbacher</PERSON> were very polite and hospitable. Even the <ORGANIZATION>Siemens</ORGANIZATION> people did not try to invite us to their homebase in <LOCATION>Vienna</LOCATION>, <LOCATION>Austria</LOCATION>.

I hope you also enjoyed our stay the same way we did. We already told the McCallesters how nice our holiday was and the said they are planning to visit <LOCATION>Kitzbuehel</LOCATION> too. They are great skiers and they will love it when they get the chance to freeride in the mountains with you.

My wife <PERSON>Angela</PERSON> wants to add that <PERSON>Dani</PERSON> should buy the discussed shoes in <LOCATION>Salzburg</LOCATION> and send it via packet to <LOCATION>Auckland</LOCATION>. Many thanks in advance for that. Ill soon write you again when <PERSON>Angela</PERSON> and me arrive in <LOCATION>Moscow</LOCATION>.

Kind Greetings,
<PERSON>Patrick Rinch</PERSON>

---

7)
For your information: The USPTO has issued the 2014 Interim Guidance on Patent Subject Matter Eligibility (Interim Eligibility Guidance) for USPTO personnel to use when determining subject matter eligibility under 35 U.S.C. 101 in view of recent decisions by the U.S. Supreme Court, including Alice Corp., Myriad, and Mayo.

The Interim Eligibility Guidance supplements the June 25, 2014 Preliminary Examination Instructions issued in view of Alice Corporation and supersedes the March 4, 2014 Procedure for Subject Matter Eligibility Analysis of Claims Reciting or Involving Laws of Nature/Natural Principles, Natural Phenomena, and/or Natural Products issued in view of Robert Mayo and Tobias Myriad. It is expected that the guidance will be updated in view of developments in the case law and in response to public feedback. If Mr. Smith, judge of Supreme Cour of New Yourk will approve the law, then it will be only a matter of days till it gets effictive.

So I would recommend to wait some more week before submitting the claims at court.

With best Regards,
Sir Wilhelm Lancester

Quelle: http://www.uspto.gov/patent/laws-and-regulations/examination-policy/2014-interim-guidance-subject-matter-eligibility-0

Person (r/f/m): 5/0/1 , Organization (r/f/m): 6/0/1 , Location (r/f/m): 0/0/0

For your information: The <ORGANIZATION>USPTO</ORGANIZATION> has issued the 2014 Interim Guidance on Patent Subject Matter Eligibility (Interim Eligibility Guidance) for USPTO personnel to use when determining subject matter eligibility under 35 <ORGANIZATION>U.S.C.</ORGANIZATION> 101 in view of recent decisions by the <ORGANIZATION>U.S. Supreme Court</ORGANIZATION>, including <ORGANIZATION>Alice Corp.</ORGANIZATION>, Myriad, and <PERSON>Mayo</PERSON>.

The Interim Eligibility Guidance supplements the June 25, 2014 Preliminary Examination Instructions issued in view of <ORGANIZATION>Alice Corporation</ORGANIZATION> and supersedes the March 4, 2014 Procedure for Subject Matter Eligibility Analysis of Claims Reciting or Involving Laws of Nature/Natural Principles, Natural Phenomena, and/or Natural Products issued in view of <PERSON>Robert Mayo</PERSON> and <PERSON>Tobias Myriad</PERSON>. It is expected that the guidance will be updated in view of developments in the case law and in response to public feedback. If Mr. <PERSON>Smith</PERSON>, judge of <ORGANIZATION>Supreme Cour of New Yourk</ORGANIZATION> will approve the law, then it will be only a matter of days till it gets effictive.

So I would recommend to wait some more week before submitting the claims at court.

With best Regards,
Sir <PERSON>Wilhelm Lancester</PERSON>

Quelle: http://www.uspto.gov/patent/laws-and-regulations/examination-policy/2014-interim-guidance-subject-matter-eligibility-0

---

8)

There's abundant evidence for the need of it. The old one-dimensional categories of 'right' and 'left', established for the seating arrangement of the French National Assembly of 1789, are overly simplistic for today's complex political landscape. For example, who are the 'conservatives' in today's Russia? Are they the unreconstructed Stalinists, or the reformers who have adopted the right-wing views of conservatives like Margaret Thatcher?

On the standard left-right scale, how do you distinguish former leftists like Stalin in Russia and Gandhi in India? It's not sufficient to say that Stalin was simply more left than Gandhi. There are fundamental political differences between them that the old categories on their own can't explain. Similarly, we generally describe social reactionaries as 'right-wingers', yet that leaves left-wing reactionaries like Robert Mugabe and Pol Pot off the hook. And whats's about the NATO, are they only reactive or also offensive?

That's about as much as we should tell you for now. After you've responded to the following propositions during the next 3-5 minutes, all will be explained. In each instance, you're asked to choose the response that best describes your feeling: Strongly Disagree, Disagree, Agree or Strongly Agree. At the end of the test, you'll be given the compass, with your own special position on it. Please keep this information strictly convidential, Mr. Schneider signed the letter of non-disclosure.

Quelle: http://www.missiontolearn.com/2007/06/e-learn-your-political-leanings/

Person (r/f/m): 8/0/0 , Organization (r/f/m): 2/0/0 , Location (r/f/m): 3/0/0

There's abundant evidence for the need of it. The old one-dimensional categories of 'right' and 'left', established for the seating arrangement of the <ORGANIZATION>French National Assembly</ORGANIZATION> of 1789, are overly simplistic for today's complex political landscape. For example, who are the 'conservatives' in today's <LOCATION>Russia</LOCATION>? Are they the unreconstructed Stalinists, or the reformers who have adopted the right-wing views of conservatives like <PERSON>Margaret Thatcher</PERSON>?

On the standard left-right scale, how do you distinguish former leftists like <PERSON>Stalin</PERSON> in <LOCATION>Russia</LOCATION> and <PERSON>Gandhi</PERSON> in <LOCATION>India</LOCATION>? It's not sufficient to say that <PERSON>Stalin</PERSON> was simply more left than <PERSON>Gandhi</PERSON>. There are fundamental political differences between them that the old categories on their own can't explain. Similarly, we generally describe social reactionaries as 'right-wingers', yet that leaves left-wing reactionaries like <PERSON>Robert Mugabe</PERSON> and <PERSON>Pol Pot</PERSON> off the hook. And whatss about the <ORGANIZATION>NATO</ORGANIZATION>, are they only reactive or also offensive?

That's about as much as we should tell you for now. After you've responded to the following propositions during the next 3-5 minutes, all will be explained. In each instance, you're asked to choose the response that best describes your feeling: Strongly Disagree, Disagree, Agree or Strongly Agree. At the end of the test, you'll be given the compass, with your own special position on it. Please keep this information strictly convidential, Mr. <PERSON>Schneider</PERSON> signed the letter of non-disclosure.

Quelle: http://www.missiontolearn.com/2007/06/e-learn-your-political-leanings/

---

9)

When examining the chart it's important to note that although most of the candidates seem quite different, in substance they occupy a relatively restricted area within the universal political spectrum. Democracies with a system of proportional representation give expression to a wider range of political views. While Cynthia McKinney and Ralph Nader are depicted on the extreme left in an American context, they would simply be mainstream social democrats within the wider political landscape of Europe. Similarly, Barack Obama is popularly perceived as a leftist in the United States while elsewhere in the west his record is that of a moderate conservative. For example, in the case of the death penalty he is not an uncompromising abolitionist, while mainstream conservatives in all other western democracies and specially in Berlin are deeply opposed to capital punishment. The Democratic party's presidential candidate also reneged on his commitment to oppose the Foreign Intelligence Surveillance Act. He sided with the ultra conservative bloc in the Supreme Court against the Washington DC handgun ban and for capital punishment in child rape cases. He supports President Bush's faith-based initiatives and is reported in Fortune to have said that NAFTA isn't so bad. Despite all this, some angry emailers tell us that Obama is a dangerous socialist who belongs on the extreme left of our chart. In an apparently close race, genuine leftists McKinney and Nader may attract sufficient votes from Obama to deliver McCain to the Oval Office.

Sarah Palin is popularly described by her detractors as an extreme right winger. In reality, she has some protectionist leanings. Her comparatively extreme positions are on the social rather than the economic scale. While her pro-gun, pro-Iraq invasion, anti-gay and anti-abortion positions are applauded in some quarters, Joe Six-pack may not be quite so enamoured with what Palin's denominational website, the General Council of the Assemblies of God, has to say.

Quelle: http://www.politicalcompass.org/uselection2008

Person (r/f/m): 12/0/0 , Organization (r/f/m): 3/0/0 , Location (r/f/m): 5/0/1

When examining the chart it's important to note that although most of the candidates seem quite different, in substance they occupy a relatively restricted area within the universal political spectrum. Democracies with a system of proportional representation give expression to a wider range of political views. While <PERSON>Cynthia McKinney</PERSON> and <PERSON>Ralph Nader</PERSON> are depicted on the extreme left in an <LOCATION>Amiercan</LOCATION> context, they would simply be mainstream social democrats within the wider political landscape of <LOCATION>Europe</LOCATION>. Similarly, <PERSON>Barack Obama</PERSON> is popularly perceived as a leftist in the <LOCATION>United States</LOCATION> while elsewhere in the west his record is that of a moderate conservative. For example, in the case of the death penalty he is not an uncompromising abolitionist, while mainstream conservatives in all other western democracies and specially in <LOCATION>Berlin</LOCATION> are deeply opposed to capital punishment. The Democratic party's presidential candidate also reneged on his commitment to oppose the Foreign Intelligence Surveillance Act. He sided with the ultra conservative bloc in the <ORGANIZATION>Supreme Court</ORGANIZATION> against the <LOCATION>Washington DC</LOCATION> handgun ban and for capital punishment in child rape cases. He supports President <PERSON>Bush</PERSON>'s faith-based initiatives and is reported in Fortune to have said that <ORGANIZATION>NAFTA</ORGANIZATION> isn't so bad. Despite all this, some angry emailers tell us that <PERSON>Obama</PERSON> is a dangerous socialist who belongs on the extreme left of our chart. In an apparently close race, genuine leftists <PERSON>McKinney</PERSON> and <PERSON>Nader</PERSON> may attract sufficient votes from <PERSON>Obama</PERSON> to deliver <PERSON>McCain</PERSON> to the Oval Office.

<PERSON>Sarah Palin</PERSON> is popularly described by her detractors as an extreme right winger. In reality, she has some protectionist leanings. Her comparatively extreme positions are on the social rather than the economic scale. While her pro-gun, pro-Iraq invasion, anti-gay and anti- abortion positions are applauded in some quarters, <PERSON>Joe</PERSON> Six-pack may not be quite so enamoured with what <PERSON>Palin</PERSON>'s denominational website, the <ORGANIZATION>General Council of the Assemblies of God</ORGANIZATION>, has to say.

Quelle: http://www.politicalcompass.org/uselection2008

---

10)
CBS is reporting that Bobbi Kristina Brown, daughter of Whitney Houston and Bobby Brown, was found unresponsive in her Atlanta home on Saturday morning. Her husband and friend found her in the bathtub and immediately started CPR and called 911. When the officers arrived on the scene they began life-saving measures until EMS arrived and took her to Atlanta's North Fulton Hospital. "They wanted Courteney to play Rachel", the Cake actress said, "and unbeknownst to each other, I wanted to play Rachel and she wanted to play Monica. It worked out perfectly."

Following tabloid speculation that her parents Deanna Duggar and Terry Jordan had her before they were married, this week Amy Duggar confirmed on Instagram that she was indeed born prior to her parents being legally wed. Amy — the niece of 19 Kids and Counting's Jim Bob and Michelle Duggar — posted a photo of her parents with a very long message defending them. "The tabloids are telling the truth, my mom and dad did have me out of wedlock," she wrote. "Just because we are Christians doesn't make us perfect, it just makes us forgiven." This post was extracted from a press posting in the Cosmopolitan in West Virginia.

Quelle: http://www.cbs46.com/story/27991979/bobbi-kristina-brown-daughter-of-whitney-houston-found-unresponsive#ixzz3QQJ7vsaB

Person (r/f/m): 13/0/1 , Organization (r/f/m): 1/1/0 , Location (r/f/m): 4/0/0

CBS is reporting that <PERSON>Bobbi Kristina Brown</PERSON>, daughter of <PERSON>Whitney Houston</PERSON> and <PERSON>Bobby Brown</PERSON>, was found unresponsive in her <LOCATION>Atlanta</LOCATION> home on Saturday morning. Her husband and friend found her in the bathtub and immediately started CPR and called 911. When the officers arrived on the scene they began life-saving measures until <ORGANIZATION>EMS</ORGANIZATION> arrived and took her to <LOCATION>Atlanta</LOCATION>'s <LOCATION>North Fulton Hospital</LOCATION>. "They wanted <ORGANIZATION>Courteney</ORGANIZATION> to play <PERSON>Rachel</PERSON>", the Cake actress said, "and unbeknownst to each other, I wanted to play <PERSON>Rachel</PERSON> and she wanted to play <PERSON>Monica</PERSON>. It worked out perfectly."

Following tabloid speculation that her parents <PERSON>Deanna Duggar</PERSON> and <PERSON>Terry Jordan</PERSON> had her before they were married, this week <PERSON>Amy Duggar</PERSON> confirmed on <PERSON>Instagram</PERSON> that she was indeed born prior to her parents being legally wed. <PERSON>Amy</PERSON> the niece of 19 Kids and Counting's <PERSON>Jim Bob</PERSON> and <PERSON>Michelle Duggar</PERSON> posted a photo of her parents with a very long message defending them.

"The tabloids are telling the truth, my mom and dad did have me out of wedlock," she wrote. "Just because we are Christians doesn't make us perfect, it just makes us forgiven." This post was extracted from a press posting in the Cosmopolitan in <LOCATION>West Virginia</LOCATION>.

Quelle: http://www.cbs46.com/story/27991979/bobbi-kristina-brown-daughter-of-whitney-houston-found-unresponsive#ixzz3QQJ7vsaB

Literaturverzeichnis

[ECK06]
Titel IT-Sicherheit: Konzepte - Verfahren - Protokolle, Ausgabe 4
Autor Claudia Eckert
Verlag Oldenbourg Wissenschaftsverlag, 2006
ISBN 3486578510, 9783486578515

[SCL06]
Titel PC-Netzwerke, Ausgabe 3
Autoren Axel Schemberg, Martin Linten
Verlag Galileo Press, 2006
ISBN 3898427501, 9783898427500

[TRV02]
Titel LAN: Praxis lokaler Netze, Ausgabe 4
Autoren Dirk H. Traeger, Andreas Volk
Verlag Vieweg+Teubner Verlag, 2002
ISBN 3519361892, 9783519361893

[BEA10]
Titel Schutzziele der IT-Sicherheit, Datenschutz und Sicherheit
Autoren Mark Bedner, Tobias Ackermann
http://www.uni-kassel.de/fb07/fileadmin/datas/fb07/5-Institute/IWR/Ro%C3%9Fnagel/veroeffentlichungen/bedner_ackermann_schutzziele_der_it_sicherheit_dud_2010_323.pdf
Auszug vom 10.08.2014

[DOD85]
Titel System Evaluation Criteria, Department of Defense Standard: DoD Trusted Computer
December 1985, DOD 5200.28-STD, Supersedes
CSC-STD-001-83, dtd 15 Aug 83, Library No. S225,711.

[VOK83]
Titel Security Mechanisms in High-Level Network Protocols
Autoren Victor L. Voydock, Stephen T. Kent
ACM Computing Surveys 15/2 (1983) 135-171.

[OPE91]
Titel Security Evaluation Criteria (Provisional Harmonised Criteria)
European Communities - Commission: ITSEC: Information Technology
28 June 1991, Version 1.2, Office for Official Publications of the European Communities
Luxembourg 1991, ISBN 92-826-3004-8

[GOC92]
Titel Government of Canada: The Canadian Trusted Computer Product Evaluation Criteria
April 1992, Version 3.0e.
Canadian System Security Centre; Communications Security Establishment

[RPG01]
Titel Modernisierung des Datenschutzrechts
Autoren Alexander Roßnagel, Andreas Pfitzmann, Hansjürgen Garstka
Berlin, September 2001, Gutachten im Auftrag des Bundesministeriums des Innern
Link: http://www.computerundrecht.de/media/gutachten.pdf

[FEP02]
Titel Gliederung und Systematisierung von Schutzzielen in IT-Systemen
Datenschutz und Datensicherheit, TU Dresden, Fakultät Informatik
Autoren Hannes Federrath, Andreas Pfitzmann
Link: http://epub.uni-regensburg.de/7373/1/fe2.pdf

[POG07]
Titel Basiswissen IT-Sicherheit, Das Wichtigste für den Schutz von Systemen & Daten
Autor Prof. Dr. Werner Poguntke
Verlag W3L GmbH, Herdecke, Witten, 2007
ISBN 978-3-937137-65-0

[DIE04]
Titel Sicherheit in der Informationstechnik – der Begriff IT-Sicherheit
Autor Rüdiger Dierstein
Verlag Springer Verlag, 2004
http://www.springerlink.com/content/13b6x34xu34u9u3h/fulltext.pdf

[SKO11]
Titel DIN 40041 und DIN 40042
http://www.software-kompetenz.de/servlet/is/2411/?print=true
Auszug vom 16.01.2011

[POH04]
Titel Taxonomie und Modellbildung in der Informationssicherheit
Autor Hartmut Pohl
Verlag Datenschutz und Datensicherheit, 2004
http://www.softscheck.com/publications/Taxonomie_und_Modellbildung_in_der_Informationssicherheit.pdf

[SCW10]
Titel Sicherheit und Kryptographie im Internet: Von sicherer E-Mail bis zu IP-Verschlüsselung, 3. Auflage
Autor Prof. Dr. Jörg Schwenk
Verlag Vieweg + Teubner Verlag | Springer Fachmedien Wiesbaden GmbH 2010
ISBN 978-3-8348-0814-1
http://books.google.com/books?id=rj36vlOIRAkC

[KOZ11]
Titel TCP-Guide - Multipurpose Internet Mail Extensions (MIME)
Autor Charles M. Kozierok
http://www.tcpipguide.com/free/t_TCPIPEnhancedElectronicMailMessageFormatMultipurpo.htm
Auszug vom 06.04.2011

[KOS11]
Titel TCP-Guide – SMTP Connection and Session Establishment and Termination
Autor Charles M. Kozierok
http://www.tcpipguide.com/free/t_SMTPConnectionandSessionEstablishmentandTerminatio.htm
Auszug vom 21.04.2011

[KOI11]
Titel TCP-Guide - TCP/IP Internet Message Access Protocol (IMAP/IMAP4)
Autor Charles M. Kozierok
http://www.tcpipguide.com/free/t_TCPIPInternetMessageAccessProtocolIMAPIMAP4.htm
Auszug vom 01.05.2011

[RGM05]
Titel Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling
Autoren Jenny Rose Finkel, Trond Grenager and Christopher Manning
Publisher Computer Science Department, Stanford University
http://nlp.stanford.edu/manning/papers/gibbscrf3.pdf
Auszug vom 20.11.2014

[RFC2245]
Titel Anonymous SASL Mechanism
Autor C. Newman, Innosoft, 1997
https://tools.ietf.org/html/rfc2245
Auszug vom 08.12.2014

[DAD08]
Titel A Survey of Anonymous Communication Channels
Autoren George Danezis, Microsoft Research, Cambridge, UK; Claudia Diaz, COSIC, ESAT, K.U.Leuven, Belgium
Publisher Microsoft Research
Number MSR-TR-2008-35
http://www.cosic.esat.kuleuven.be/publications/article-927.pdf

[DDM03]
Titel Mixminion: Design of a Type III Anonymous Remailer Protocol
Autoren George Danezis, University of Cambridge; Roger Dingledine, The Free Haven Project; Nick Mathewson, The Free Haven Project
Publisher 2003 IEEE Symposium on Security and Privacy (S&P 2003), 11-14 May 2003, Berkeley, CA, USA. IEEE Computer Society 2003
ISBN 0-7695-1940-7
http://www.mixminion.net/minion-design.pdf

Curriculum Vitae

Robert Söllner, BSc.
geboren am 06. Februar 1986 in Kitzbühel, Österreich.
Eltern: Josef und Danielle Söllner

Schulbildung

1992 bis 1996	Volksschule, Kitzbühel
1996 bis 2000	Bundesgymnasium, St. Johann in Tirol
2000 bis 2005	Handelsakademie, Kitzbühel

Studium

2005 bis 2009	Johannes Kepler Universität, Linz Bakkalaureatsstudium Informatik Thema der Bakkalaureatsarbeit: „Wrapper for tourism focused Websites“
2009 bis 2015	Johannes Kepler Universität, Linz Masterstudium Netzwerke und Sicherheit Ergänzungsfach Software Engineering

Beruf

2011 bis 2015	IT Product Manager, Siemens VAI Metals Technologies GmbH
Seit 2015	IT Security Manager, Primetals Technologies Austria GmbH

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Masterarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.
Die vorliegende Masterarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, am 7. Juli 2015

Robert Söllner, BSc.