Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Bachelorarbeit

Entwicklung eines Visualisierungswerkzeuges zur Demonstration datenschutzfreundlicher Dokumentspeicherdienste

vorgelegt von

David Kirchhausen Monteiro

geb. am 24. Januar 1994 in Hildesheim

Matrikelnummer 6530927

Studiengang Software-System-Entwicklung

eingereicht am 8. Juni 2018

Betreuer: Maximilian Blochberger

Erstgutachter: Prof. Dr.-Ing. Hannes Federrath

Zweitgutachter: Tilmann Stehle, M. Sc.

Aufgabenstellung

Im Zuge dieser Bachelorarbeit soll ein einfacher Dokumentenspeicher entwickelt werden, welcher mögichst viele Nutzerdaten erfasst und speichert. Die erfassten Daten sollen anschaulich grafisch dargestellt werden können. Weiter sollen verschiedene Szenarien entwickelt werden, welche aufzeigen wie eine mögliche Benutzung des Services mit und ohne der Verwendung von datenschutzfreundlichen Methoden zum Anonymisieren von Daten aussieht. Anhand der Szenarien soll eine grafische Auswertung Unterschiede zwischen anonymisierten Daten und nicht anonymisierten Daten visuell sichtbar machen und die Unterschiede somit leicht zugänglich sein.

Zusammenfassung

1. Dokumentenspeicherdienste Vorteile (Problemstellung erläutern)
2. Mögliche Datenschutz unfreundliche Aspekte von gängigen Anbietern (Problemstellung erläutern)
3. Entwicklung des Dokumentenspeichers und der Visualisierung zur deutlich Veranschaulichung von Potentiellen Unterschieden zwischen der Verwendung von Datenschutz freundlichen Methoden zum Anonymisieren oder nicht. (Bearbeitung der Problemstellung)
    a) Implementation des Dokumentenspeichers
    b) Implementation der API zur Datenübergabe
    c) Implementation des Visualisierungswerkzeug
    d) Darstellung der Szenarien zur Benutzung des Visualisierungswerkzeug

Inhaltsverzeichnis

1 Einleitung **5**
1.1 Problemstellung 5
1.2 Problembearbeitung 5

2 Hauptteil **6**
2.1 Grundlage: Terminologie 6
2.2 Grundlage: Set A / Set B 6
2.3 Darstellung : IP Tree Map und IP Google Map 6
2.4 Darstellung: Headerfingerprinting 7
2.5 Darstelung: Time Line 7

3 Schluss **8**
3.1 Zusammenfassung der Ergebnisse 8
3.2 kritische Bewertung des Ergebnisse 8
3.3 neue Problemstellungen, Möglichkeiten zur Weiterführung der Arbeit 8

1 Einleitung

1.1 Problemstellung

1. Dokumentenspeicherdienste biten vorteile und Nachteile
2. möglichen Datenschutzmissbrauch durch Anbieter aufzeigen
3. Mögliche Verfahren/Methoden zum Datenschutz des Klienten aufzeigen
4. Implementation des Dokumentenspeicherdienstes als negativ Beispiel.
5. Nutzung des Dokumentenspeichers zur Gegenüberstellung von Datenschutz freundlichen und Datenschutz unfreundlichen Methoden zum Anonymisieren von Daten
6. Gewinn an Visuellen klar erkennbaren Unterschieden für z.B. nicht Informatiker zum besseren Verständnis des Problems

1.2 Problembearbeitung

1. Implementation des Dokumentenspeichers
    a) Vorstellung des Dotnet Core Framework
    b) Implementierung der Api des Dokumentenspeichers
        i. Api für die Verwendung des Dienstes
        ii. Api für die Ausgabe der Relevanten Daten
    c) Vorstellung des D3.js Framework zur Visualisierung
    d) Vorstellung der verschiedenen Komponenten zur Visualisierung
2. Schematische Verwendung des Dokumentenspeichers für die Problembearbeitung

2 Hauptteil

2.1 Grundlage: Terminologie

Einführung von Begriffen zur Beschreibung von Teil Aspekten des Dokumentenspeichers und der Verwendung

2.2 Grundlage: Set A / Set B

Das Visualisierungswerkzeug besitzt zwei verschiedene Datenbanken welche Set A und Set B genannt werden. Jedes Datenbank Set verfügt über eine eigene API und funktioniert identisch. Die angestrebte Benutzung sieht den Vergleich von Set A und Set B mit den gleichen Datensatz vor wobei die nicht Anwendung von Datenschutz freundlichen Methoden zum Anonymisieren bei Set A und die Anwendung von Datenschutz freundlichen Methoden zum Anonymisieren bei Set B. Set A ist somit die Basis und zeigt auf was ein Dokumentenspeicherdienst an Nutzerdaten sammeln kann und Set B kann im direkten Vergleich zeigen wie Datenschutz freundlichen Methoden zum Anonymisieren diese Daten verfälschen.

2.3 Darstellung : IP Tree Map und IP Google Map

Die Darstellung der gesammelten Daten über IP-Adressen Gruppierte Mengen.

Erzeugen des Gruppierungen aus der gegebenen Datenmenge. Nutzung des d3.js zur Darstellung der Tree Map. Dabei wird jeder IP-Gruppierung eine andere Farbe zugeordnet. Die Farbliche Visuelle Darstellung macht zusammen gehörige Dateien nach IP-Adresse direkt sichtbar. Mit Hilfe der IP-Adressen und eines Geolookup kann eine ungefährer Standpunkt (Lat/Lon) der IP-Adresse gewonnen werden. Anhand dieses Standpunkts kann auf der Google Map der Standpunkt von wo eine Datei hochgeladen wurde aufgezeigt werden.

Bei Verwendung von Methoden zur Anonymisieren der IP-Adresse wie z.B. das verwenden eines Proxys oder der Verwendung des TOR-Netzwerks, werden die IP-Adressen-Gruppierungen verzerrt, durch verzerrte Gruppen können Dateien eines Benutzer nicht mehr zu 100% auf diesem Benutzer gemappt werden.

1. Proxy -> IP-Adresse wird maskiert
    a) Falls der Proxy wird nur durch einen Benutzer benutzt -> keine Gruppenverzerrung
    b) Proxy wird von mehreren Benutzern benutzt -> Gruppenverzerrung
    c) Benutzer wechselt Proxy mehrfach -> Gruppierungen werden in der Gesamtheit verzerrt -> pro Benutzer mehre Gruppierungen

2. IP-Gruppierung über die Endknoten des Tor-Netzwerks
    a) Benutzer die den gleichen Endknoten benutzen werden gruppiert -> Gruppenverzerrung
    b) Durch automatischen wechsle der Endknoten -> Pro Benutzer zwangsläufig mehrere Gruppen -> Gruppenverzerrung

2.4 Darstellung: Headerfingerprinting

2.5 Darstelung: Time Line

3 Schluss

3.1 Zusammenfassung der Ergebnisse

3.2 kritische Bewertung des Ergebnisse

3.3 neue Problemstellungen, Möglichkeiten zur Weiterführung der Arbeit