HSR
HOCHSCHULE FÜR TECHNIK
RAPPERSWIL

FHO Fachhochschule Ostschweiz

Bachelorarbeit, Abteilung Informatik

Background Traffic Monitoring

Hochschule für Technik Rapperswil

Frühlingssemester 2019

Autoren:	Andi Hörler und Jonas Kugler
Betreuer:	Prof. Dr. Peter Heinzmann
Externe Ansprechpartner:	Klaus Degner (Allegro Packets), Eric Franke (cnlab itr AG)
Arbeitsperiode:	18.02.2019 - 14.06.2019
Arbeitsumfang:	360 Stunden, 12 ECTS pro Student

Ein spezieller Dank für die Unterstützung, das Korrekturlesen und die Teilnahme an Usability Tests geht an folgende Personen:

Eric Franke

Felix Kugler
Tobias Saladin

Inhaltsverzeichnis

Abstract

Mit der wachsenden Anzahl von ans Internet angeschlossenen Geräten ging sowohl in Firmennetzen als auch im Privatbereich die Übersicht verloren, welche Geräte aktiv sind und zu welchen externen Stellen Daten übertragen werden. Viele Internet Nutzer sind sich nicht bewusst, dass beispielsweise beim Aufruf von www.20min.ch ohne Zustimmung des Endnutzers im Hintergrund weitere Verbindungen zu anderen Endpunkten aufgebaut werden.
Es existieren bereits einige Tools für die Auswertung von Netzwerkverkehr, doch ist deren Bedienung kompliziert und erfordert entsprechendes Know-How.

In dieser Bachelorarbeit soll eine Applikation entwickelt werden, welche sowohl interessierten Laien als auch erfahrenen Informatikern Antworten zu folgenden Fragen in Bezug auf Hintergrundverkehr liefert:

- Welche weitere Stellen werden beim Aufruf einer bestimmten Webseite noch kontaktiert?
- Welche Geräte sind in einem Netz aktiv?

Zusätzlich soll aufgezeigt werden, ob veraltete Security Protokolle zum Einsatz kommen. Die Applikation soll im Rahmen von Schulungen und zur Analyse der Situation bei Heim- und Firmennetzwerken genutzt werden können. Aus Aufzeichnungen des Netzverkehrs sollen Informationen zum Hintergrundverkehr extrahiert, interpretiert und in grafischer und tabellarischer Form präsentiert werden, um das Bewusstsein über Hintergrundverkehr zu steigern.

Die entwickelte App, für die Analysesoftware Splunk, reichert den aufgezeichneten Netzwerkverkehr mit Zusatzinformationen an. Mithilfe der App ist eine vereinfachte Analyse von Hintergrundverkehr möglich.
Beispielsweise wurde für einige in der Schweiz verbreitete Webseiten aufgezeigt, welche zusätzlichen Endpunkte angesprochen werden. Dabei zeigte sich, dass die Datenschutzerklärungen nicht immer genau angeben, welche zusätzliche Endpunkte verwendet werden.
Die Analyse von drei Speedtests für Mobile Devices ergab, dass bei Apps mit sehr ähnlichem Funktionsumfang grosse Unterschiede in der Anzahl und Art der im Hintergrund

zusätzlich aufgebauten Netzwerkverbindungen bestehen: Der Speedtest von cnlab baut 37, der von OpenSignal 87 und jener von Ookla 164 Verbindungen auf.
Eine Analyse des Verhaltens bei Webseitenaufrufen mit und ohne Adblocker zeigt, dass ohne Adblocker wesentlich grössere Datenmengen übertragen werden als mit Adblocker: 40% mehr ankommende und 352% mehr ausgehende Daten. Bei der Adblocker Analyse konnte auch gezeigt werden, dass ohne Adblocker mehr veraltete Chiffren zum Einsatz kommen.
Die App wird in den kommenden Wochen im Splunk Appstore, der Splunkbase, veröffentlicht.

Aufgabenstellung

HSR
HOCHSCHULE FÜR TECHNIK
RAPPERSWIL

FHO Fachhochschule Ostschweiz

Background Traffic Monitoring

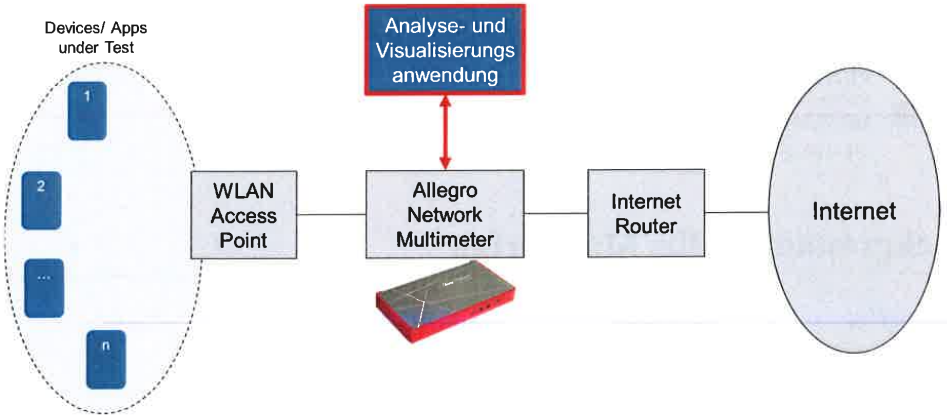
Studiengang:	Informatik
Semester:	FS 2019
Durchführung:	Bachelorarbeit
Fachrichtung:	Internet-Technologien und -Anwendungen
Institut:	INS
Studenten:	Andi Hörler und Jonas Kugler
Betreuer:	Peter Heinzmann (HSR/cnlab)
Externe Ansprechpartner:	Klaus Degner (Allegro Packets), Eric Franke (cnlab)

Ausgangslage

Rechner, Tablets, Smartphones und Unterhaltungselektronikgeräte sind mittlerweile dauernd mit dem Internet verbunden. Für normale Nutzer ist nicht ersichtlich, welche Daten mit welchen Endstellen im Hintergrund ausgetauscht werden. Die Art der Hintergrunddatenübertragung reicht von automatisierten Updates, über verschiedenste Werbetracker bis hin zu verstecktem Bildschirm-Monitoring.

Für die Situationsanalyse wäre ein Werkzeug nützlich, welches alle Verbindungen aufzeigt und klassifiziert. Für jeden aufgerufenen Server und für jeden Rechner oder für jede App würde man sehen, mit welchen Endpunkten wie viele Daten ausgetauscht werden. Dazu müssen Metainformationen zu den Verbindungen generiert werden, welche beispielsweise anzeigen, ob es sich um Tracking Server, Advertising Server oder direkt angesprochene Server handelt. Diese Metainformationen könnten auch anzeigen, von welchen Servern Videoinhalte, Audioinhalte, Bilder, Texte oder andere Daten geliefert werden.

Im Rahmen dieser Bachelorarbeit soll ein System entwickelt werden, mit welchem Nutzern aufgezeigt werden kann, was alles im Hintergrund abläuft, wenn ein Rechner oder Smartphone mit dem Internet verbunden ist. Es soll ersichtlich sein, was im passiven Zustand und was bei der Nutzung bestimmter App bzw. Programme passiert.



Ziel

Für Endgeräte, welche über Allegro Packet Monitoring mit dem Internet verbunden sind, sollen alle Verbindungen, visualisiert und klassiert werden können.

Unter anderem sind folgende Informationen anzuzeigen:

1. Welche **Sites** werden **von einem bestimmten Endgerät** während der Nutzung (z.B. Surfen oder YouTube schauen) und im unbenutzten Zustand (z.B. Smartphone im Standby) aufgerufen?
    - Anzahl der Sites (Domains)
    - Art des Aufrufs (http, https, QUIC, ...)
    - Hoch- und heruntergeladene Datenmenge
    - Logische Anzeige der Sites (vgl. Lightbeam, DisconnectMe)
        - Rechner in einem Subnetz
        - Rechner, welche innerhalb einer bestimmten Zeitperiode aufgerufen wurden
        - Rechner mit verschiedenen Aufgaben (Tracker, CDN, DNS)
        - Rechner mit bestimmten verwendeten Verschlüsselungsverfahren
    - Geografische Anzeige der Sites
        - Kann Allegro die Ortsauflösung liefern?
        - Cnlab verwendet aktuell die GeoIP-API von Google
        - Gibt es bessere GeoIP-APIs
    - Resultat
        - Auswertung für bestimmte Endgeräte (z.B. privater Rechner, Smartphone, Digitale Assistenten wie Alexa oder Google Home)
        - Detektion von verdächtigen Sites, welche "ungewöhnlich" grosse Datenmengen erhalten bzw. zu denen bestimmte Clients viele Daten senden.
            - Völlig unbekannte Sites sollten eher keine grossen Datenmengen erhalten, Google Photos darf durchaus grosse Datenmengen erhalten
            - Für die Bestimmung von Detektionsverfahren soll in der Einarbeitungsphase Verkehr mit Allegro beobachtet werden.
        - Klassierung der Sites (z.B. Advertising, Tracking, Content Distribution)
            - Basierend auf Listen (z.B. Ad-Server Listen)
            - Basierend auf Algorithmen
        - Untersuchung und Charakterisierung des Einflusses von Ad-Blockern

2. Welche **Sites und weitere URLs** werden **von bestimmten Apps und Programmen** aufgerufen?
    - Wie kann man zusammengehörige Aufrufe einer App (oder des Browsers) erkennen?
        - Zeitliche Folge
        - IP-ID
        - Referer bei http-Verkehr
    - Falls man Apps erkennen kann, so sollen dazu charakteristische Parameter angezeigt werden
        - Anzahl der URLs
        - Anzahl der Sites (Domains)
        - Art des Aufrufs (http, https, QUIC, ...)
        - Hoch- und heruntergeladene Datenmenge
        - Logische Anzeige der Sites (vgl. Lightbeam, DisconnectMe)
        - Geografische Anzeige der Sites
    - In der Einarbeitungsphase soll analysiert werden, wie Cookie Richtlinien eingehalten werden (werden wirklich keine Cookies gesetzt, wenn ein Benutzer das nicht wünscht)
        - Dies macht man vor allem auf dem Client bzw. Browser Entwickler Modus und schaut zusätzlich, was Allegro dazu anzeigt.
    - Für die Bestimmung von Detektionsverfahren soll in der Einarbeitungsphase Verkehr bei bekannten Tracking/Screenshot/Sniffing-Apps beobachtet werden (z.B. www.glassboxdigital.com oder mouseflow.com).
    - Resultat ähnlich wie oben, aber für bestimmte Apps (z.B. Aufruf von News-Diensten mit einer App und mit dem Browser, Nutzung von Google Maps, Whatsapp)
        - Detektion von "verdächtigen" Sites, welche eigentlich gar nicht aufgerufen wurden (Definition, was unter "verdächtig" zu verstehen ist.)
        - Klassierung der Sites (z.B. Advertising, Tracking, Content Distribution)
        - Untersuchung und Charakterisierung des Einflusses von Ad Blockern

Ob die Auswertungen direkt auf der Allegro Plattform oder extern angezeigt werden sollen, ist zusammen mit Allegro abzuklären. Wahrscheinlich macht es Sinn, in einer ersten Phase, die Auswertungen extern anzuzeigen.

Aufgaben

- Einarbeitung / Analyse
  - Auffinden und studieren von Berichten und ähnlichen Arbeiten zum Hintergrunddatenverkehr
    - Klassierung der verschiedenen Arten von Hintergrunddatenübertragung
    - Präzisierung der Aufgabenstellung
  - Einarbeitung und Beschreibung des Internet Advertising Ecosystems
    - Einstiegshilfe der Betreuer (Workshop, Kurs)
    - Praxistests, Verbesserung der Unterlagen
  - Allegro Einarbeitung
  - Analyse der Einhaltung von Cookie Richtlinien
  - Analyse des Datenverkehrs bei Alexa / Google Home
- Design
  - "Produktbeschreibung"
  - Auswertungsalgorithmen
  - Visualisierungsformen
- Realisierung
  - Auswerteportal
- Testing / Demonstration
  - Aufzeigen des Hintergrundverkehrs für verschiedene Anwendungsfälle (aktive und passive Situation bei Private Smartphone, Tablet Nutzung; Heimnetzumgebung mit Unterhaltungselektronik (inkl. Alexa und Google Home) und IoT Geräten, Arbeitsrechner mit iOS und Windows)
  - Aufzeigen des Hintergrundverkehrs für verschiedene Anwendungsfälle
    - aktive und passive Situation bei Private Smartphone
    - Tablet Nutzung
    - Heimnetzumgebung mit Unterhaltungselektronik wie z.B. PlayStation, Xbox, TV-Boxen
    - Digitale Assistenten wie z.B. Alexa und Google Home
    - IoT Geräte
    - Arbeitsrechner mit iOS und Windows

Referenzen, Beispiele

1. Hinweise zur Durchführung von Studienarbeiten:
   https://drive.switch.ch/index.php/s/nam69AOJabY2XhG
2. Zack Whittaker, "Many popular iPhone apps secretly record your screen without asking And there's no way a user would know", TechCrunch, 7.1.2019
   https://techcrunch.com/2019/02/06/iphone-session-replay-screenshots/
3. Allegro Network Mulitmeter https://allegro-packets.com/de/
4. Firefox Lightbeam Add-On https://addons.mozilla.org/de/firefox/addon/lightbeam
5. Disconnect Me, Werbetracker und -blocker Add-On, https://disconnect.me
6. Internet Advertising (Unterlangen Datenschutzkurse von P. Heinzmann)
   19.6.2014 How an Ad is Served with Real Time Bidding (RTB) - IAB Digital Simplified
   Learn the back-end process of how a targeted ad is served to you from your computer, through the multifaceted pipeline of the digital advertising ecosystem.
   www.youtube.com/watch?v=-Glgi9RRZJs 5m27s
   19.05.2014, Pete Kluge, Adobe, Display Advertising Basics (sehr gute Übersicht zu den Grundbegriffen) https://youtu.be/xnX1nxMM_R0 20m12s
   19.5.2017 Sacha Berlik, The Trade Desk, Truth about Programmatic (excellente Präsentation)
   www.youtube.com/watch?v=pENf93b6qAY 18m14s
   Programmatic advertising is innovating and evolving an entire industry at breakneck speed. While these advertising advancements are exciting, there's a difference between what is coming in the future and what is possible today.
7. Dominic Peisker, Matthias Fehr, "Online Advertising", Bachelorarbeit Frühjahrssemester 2016, Hochschule für Technik Rapperswil, Juni 2016.
8. Cisco, Encrypted Traffic Analytics, White Paper, January 2019
   https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/enterprise-network-security/nb-09-encrytd-traf-anlytcs-wp-cte-en.pdf?dtid=oblgzzz000659
9. Cisco Joy, Joy is a BSD-licensed libpcap-based software package for extracting data features from live network traffic or packet capture (pcap) files, using a flow-oriented model similar to that of IPFIX or Netflow, and then representing these data features in JSON. It also contains analysis tools that can be applied to these data files. Joy can be used to explore data at scale, especially security and threat-relevant data.
   https://developer.cisco.com/codeexchange/github/repo/cisco/joy/

Management Summary

Durch die wachsende Anzahl an Smart Home Devices und die aktive Nutzung von Cloud-Diensten ging die Übersicht der übermittelten Daten an externe Stellen verloren. Viele Internet Nutzer sind sich nicht bewusst, dass ohne ihre Zustimmung im Hintergrund Daten mit weiteren, nicht direkt vom Nutzer angesprochenen Servern, ausgetauscht werden.
Zwar existieren Tools, wie Wireshark, welche den Netzwerkverkehr aufzeichnen und anschliessend auswerten können, jedoch sind diese mit einem entsprechenden Knowhow zu bedienen, um an die benötigten Informationen zu gelangen.

Das Ziel dieser Bachelorarbeit ist, eine Software zu programmieren, welche anhand von aufgezeichnetem Netzwerkverkehr eine Auswertung mit vordefinierten Filtern und Darstellungen ermöglicht. Die Ansichten sollen dabei tabellarisch oder graphisch die gewünschten Informationen zu den grundlegenden Fragen aufzeigen:

- Welche weiteren Stellen werden beim Aufruf einer bestimmten Webseite noch kontaktiert?
- Welche Geräte sind in einem Netz aktiv?
- Wie sicher ist die Kommunikation zwischen zwei Endpunkten wirklich?

Die Applikation wird mithilfe der bestehenden Analysesoftware Splunk umgesetzt. Splunk bietet den notwendigen Grundbau, eine Applikation zu entwickeln, welche Rohinformationen innerhalb von Netzwerkcaptures von verschiedenen Orten verarbeitet, in ein von Splunk lesbares Format umgewandelt, um dieses dann in Splunk einzulesen. Splunk bietet auch für die Darstellungen der Daten Funktionen an, welche eine kompakte Übersicht der übermittelten Daten liefern können.

Mit der entwickelten App wurden anschliessend Analysen durchgeführt, welche den ausgearbeiteten Personas entsprechen.
So haben Tests mit und ohne Adblocker gezeigt, dass die Nutzung ohne einen Adblocker die Anzahl an Verbindungen zu potentiell unerwünschten Server wesentlich vergrössert (+ 352%). Ohne Adblocker steigt zudem die Anzahl an unsicher übermittelten Packets

an (+ 283%).
Mithilfe der App konnte zudem analysiert werden, ob die Datenschutzerklärungen einzelner Webseiten auch wirklich die Anbieter listen, mit welchen bei einem Webseitenaufruf ohne Adblocker Verbindungen aufgebaut wurden.
Neben Tests mit Webseiten wurden noch Speedtests analysiert. Dabei konnte gezeigt werden, dass es wesentliche Unterschiede in der Anzahl aufgebauter Verbindungen (von 37 bis 164), der übertragenen Datenmenge im Up- und Download und bei der Verwendung von Cipher Suites gibt.

1 Einleitung

1.1 Ausgangslage

Laut dem Bundesamt für Statistik (BFS) nutzen im Jahre 2018 90% der Schweizer Bevölkerung im Alter zwischen 16 und 74 Jahren das Internet mindestens einmal pro Woche [1]. Etwas weniger Leute, knapp über 80% nutzen das Internet sogar täglich [2, S. 3]. Zahlen aus dem Jahre 2017 zeigen zudem, dass 65% der Internetnutzer in der Schweiz fünf oder mehr Stunden pro Woche im Internet unterwegs sind [3].
Aufgrund des hohen Prozentsatzes der Schweizerinnen und Schweizer, welche das Internet nutzen, ist es auch nicht weiter verwunderlich, dass 87% der Internetnutzer personenbezogene Daten über das Internet weitergeben. Darin enthalten sind auch jene rund 50% der Schweizerinnen und Schweizer, welche laut Studie Fotos, Standortangaben und Informationen betreffend Gesundheitszustand und Beschäftigungssituation weitergeben [2, S. 7]. Etwas mehr als 80% der befragten Schweizerinnen und Schweizer gaben zudem an, dass sie Massnahmen zum Schutz von personenbezogenen Daten treffen, wobei die dabei meistgenannte Massnahme das Verweigern der Zustimmung für die Weitergabe der gesammelten Daten ist [2, S. 8].
Auch wenn sich die Schweiz bis hier im europäischen Vergleich noch gut geschlagen hat, indem sie jeweils über dem europäischen Durchschnitt angesiedelt war, wird bei etwas technischeren Fragen das Defizit der Schweiz aufgezeigt. So wussten nur 62% der Schweizerinnen und Schweizer zwischen 16 und 24 Jahren, dass ihr Tun im Internet mittels Cookies nachverfolgt werden kann. Damit liegen sie 10 Prozentpunkte hinter dem europäischen Schnitt [2, S. 10].

Zusammengefasst zeigen die Zahlen, dass die Schweizer Bevölkerung zwar gegenüber den Gefahren im Internet sensibilisiert wird, dies aber nicht genügen mit technischen Informationen zu den Gefahren verknüpft wird.

1.2 Hochschule für Technik Rapperswil

An der HSR hat Prof. Heinzmann einen Lehrgang für betriebliche Datenschutzbeauftragte aufgebaut [4]. Im Rahmen von diesem Lehrgang, aber auch an anderen Veranstaltungen soll die Möglichkeit geboten werden, dass Leute, die über einen Internet Anschluss mit Background Traffic Monitoring surfen, «am eigenen Leib» erfahren, an welche Stellen

Daten geliefert werden, wenn sie surfen.

1.3 Ziel

Es soll eine Applikation erstellt werden, welche Daten aus Netzwerk-Mitschnitten mit datenschutzrelevanten Informationen anreichert und in einer Form darstellt, welche sowohl Personen vom Fach wie auch interessierten Laien einen schnellen Überblick über die aufgezeichnete Situation geben. Somit soll sowohl die Suche nach Auffälligkeiten im Netzwerk wie auch die Präsentation der Daten für Schulungen ermöglicht werden, ohne selbst die Daten mit Tools, wie zum Beispiel Wireshark, auswerten zu müssen.

2 Grundlagen

Im Kapitel Grundlagen wird eine kurze Übersicht über Begriffe vermittelt, welche im Verlauf dieser Arbeit gebraucht werden.

2.1 Ad

2.1.1 Entwicklung

Vor genau 25 Jahren, im Jahre 1994, wurde der Grundstein für die Online Werbung gelegt. Als Teil der «You will» Werbekampagne liess AT&T auf der Webseite HotWired.com das erste Werbebanner der Geschichte aufschalten.

Abbildung 2.1: Das erste Werbebanner [5]

Anders als heute, war damals eines der grössten Probleme solcher Werbungen, dass viele Firmen, die Werbung schalten wollten, noch überhaupt keinen Webauftritt hatten, auf welchen verlinkt werden konnte. [6]
Die Entwicklung der Online-Werbung hat sich aufgrund ihres Erfolges ständig weiterentwickelt. So ging es über Graphics Interchange Formats (GIFs) und Popups weiter bis zum heutigen Stand mit Real Time Biding (RTB) und Programmatic Advertising.

Programmatic Advertising

Unter Programmatic Advertising versteht man das automatisierte Kaufen von Werbeflächen. Dabei wird dieser Prozess, verglichen mit dem manuellen Kaufvorgang, durch Algorithmen verbessert und optimiert. Bei Programmatic Advertising können auch Technologien, wie zum Beispiel das maschinelle Lernen, zum Einsatz kommen, um stark personalisierte Werbung ausliefern zu können. [7]

Real Time Biding

RTB ist eine mögliche Art Programmatic Advertising zu betreiben, wobei aktuell etwa 90% der Werbefläche auf diese Weise gekauft wird. Bei RTB werden Werbeflächen auf Webseiten oder andere Online-Werbeflächen auf einem Exchange versteigert. Sobald jemand eine Webseite besucht, welche Werbeflächen über RTB verkauft, wird die Werbefläche für diesen spezifischen User zum Verkauf angeboten. Abhängig von Informationen über den Webseitenbesucher geben interessierte Werber nun automatisiert Gebote ab. Die Werbung des Höchstbietenden wird am Ende dem Webseitenbesucher angezeigt. Dieser gesamte Prozess dauert nicht länger als 100 Millisekunden [8]. Die Zeit, bis einem User die Werbung schlussendlich angezeigt wird, kann abhängig von dessen Internetanbindung um ein Vielfaches grösser sein [9, S. 34].

2.1.2 Werbung in Zukunft

Um potenzielle Kunden noch besser mit Werbung ansprechen zu können, um sie von potentiellen zu effektiven Kunden zu machen, wird es immer wichtiger Werbung möglichst direkt auf die Person zuzuschneiden. Ein Beispiel für eine solche Art der Werbung lieferte Zalando im Jahre 2015. Damals starteten sie eine Werbekampagne, um Kunden für ein neues Angebot im europäischen Raum zu gewinnen. Dafür wurden über 60'000 Werbevideos mit kleinen Unterschieden erstellt. Das Ziel davon war, dass, egal wo in Europa die Werbung angesehen wurde, der entsprechende Ortsname im Video vorkam. [10]

2.1.3 Adblocker

Um der zunehmenden Flut an Werbungen im Internet entgegen zu wirken, haben sich auch die Adblocker über die Jahre weiterentwickelt. Während Adblocker immer populärer wurden, versuchten Entwickler hinter einigen dieser Programme mehr Gewinne daraus zu ziehen, indem sie Geld dafür verlangen, Werbungen in eine Whitelist aufzunehmen. Dies geschah beispielsweise beim Programm «Adblock Plus». Um auf einer solchen Whitelist landen zu können, müssen auch gewisse Vorgaben erfüllt werden, die sicherstellen sollen, dass die Werbung den Benutzer nicht übermässig stört [11]. Viele User störten sich daran, dass trotz installiertem Adblocker Werbung angezeigt wurde und wechselten darum auf einen anderen Adblocker.
Die aktuelle Entwicklung geht trotzdem weiter in Richtung Adblocker mit Whitelists. Google selbst wird einen Adblocker veröffentlichen, welcher darauf ausgelegt ist, nur störende Werbung zu blockieren [12]. Gleichzeitig plant Google Änderungen an der Funktionalität ihres Chrome Browser, welche von den Adblockern für das Herausfiltern von Werbung genutzt wird. Die Reduktion der Anzahl erlaubter Filterregeln würde viele dieser Adblocker in ihrer Funktion einschränken [13]. Die Zukunft von Adblockern wie sie heute existieren ist daher ungewiss.

2.2 Tracking

Als Tracker werden in Applikationen und Webseiten eingebundene Elemente bezeichnet, welche Informationen über den Benutzer und dessen Verhalten sammeln und diese Informationen anschliessend an einen eigenen Server senden. Die so gesammelten Informationen werden beispielsweise für das Anzeigen von personalisierter Werbung genutzt.
Ein Beispiel für solche Elemente sind Buttons von Sozialen Netzwerken, welche genutzt werden können, um Webseiten zu «liken». Wird eine Webseite mit einem solchen Button geladen, können die über den Benutzer gesammelt Informationen übermittelt werden. [14]

2.3 Splunk

Splunk ist eine Software, welche als Analysetool von Log-Files jeglicher Art verwendet werden kann. Splunk ist dabei seit mehreren Jahren führender Anbieter von Security Information and Event Management (SIEM) Software [15].
Splunk kann mit einer Three-Tier Architektur verglichen werden. Durch die Installation von Splunk wird ein Backend, eine Datenbank und ein Frontend eingerichtet.

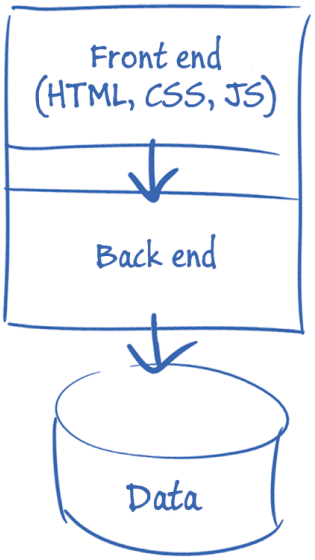


Abbildung 2.2: Splunk Basis Architektur [16]

Die Datenbank dient bei der Architektur als persistente Speicherung von Daten. Splunk verwendet dabei eine eigene Datenbankstruktur, welche einzulesende Daten konvertiert und als Events [17] ablegt.
Mithilfe vom Splunk Frontend können Daten von der Datenbank abgefragt und visualisiert werden. Splunk setzt dabei auf eine eigene Abfragesprache Search Processing

Language (SPL), welche eine Kombination von Pipes von Unix und Abfrage Statements von Structured Query Language (SQL) aufbaut [18]. Die Abfragen und Visualisierungen können dabei in Dashboards abgespeichert werden.
Das Backend von Splunk ermöglicht es Skripts auf dem Splunk Server auszuführen. Dabei wird keine Skriptsprache durch Splunk vordefiniert. Das Backend ermöglicht es auch Ordner oder Dateien zu monitoren, zu verarbeiten und diese anschliessend als Events in die Datenbank einzulesen. Splunk nutzt hierzu Konfigurationsdateien, welche Prozesse definieren. Das Backend liefert zudem noch eine Command Line Interface (CLI)-Integration, mit welcher einzelne Splunk Tasks individuell gestartet werden können.

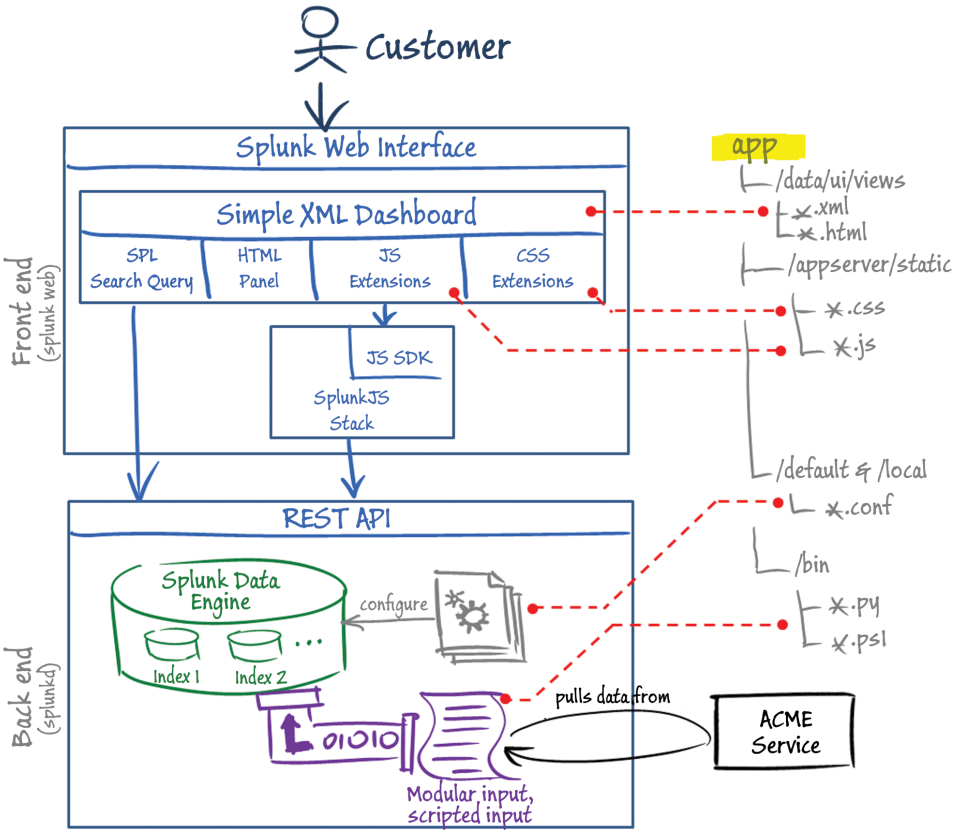


Abbildung 2.3: Splunks detaillierte Architektur [16]

2.3.1 Konfigurationsdateien

Mittels Konfigurationsdateien kann das Backend von Splunk individualisiert werden. So können beispielsweise Datentypen, Prozesseschritte oder das Einlesen von Daten definiert werden. Konfigurationsdateien besitzen dabei die Dateiendung .conf. [19]

2.3.2 Source Types

Eingelesene Dateien in Splunk können als bestimmte Source Types deklariert werden. Dies erleichtert den anschliessenden Filter Prozess. Eigene Source Types werden in der Konfigurationsdatei props.conf hinterlegt [20] und werden danach beim Einlesen der Files referenziert.

2.3.3 Stanzas

Splunk nutzt für Konfigurationsdateien sogenannte Stanzas. Stanzas definieren einzelne Bereiche von den Konfigurationsfiles. Stanzas besitzen dabei Key/Value Pairs, welche entweder über Splunk oder individuell per CLI konfiguriert werden können. [21]

2.3.4 Splunk Apps

Die zuvor erwähnten Konfigurationen der Tiers lassen sich in einer Splunk App zusammenfassen und speichern. Splunk Apps werden innerhalb der Splunk Instanz installiert. Splunk Apps können selbst erstellt werden und als Archivdateien geteilt werden. Dabei bietet das Frontend oder das CLI die Möglichkeit, diese Archivdateien als Splunk Apps auf weiteren Splunk Systemen zu installieren. Jede Splunk App orientiert sich dabei an den vorgegebenen Splunk App Richtlinien [22]. Splunk Apps können dabei Dashboards, Konfigurationen und Skripts beinhalten.

2.3.5 Splunkbase

Die Splunkbase ist ein App-Store für Splunk Apps. Um Apps von der Splunkbase herunterladen zu können, wird ein Splunk Account benötigt. Auch das Publishen von Splunk Apps benötigt einen Splunk Account. Die Splunkbase umfasst zurzeit 2044 Apps (Stand vom 19.05.2019).

2.4 Packet Caputre

Um Netzwerkverkehr analysieren zu können, muss dieser aufgezeichnet und abgespeichert werden können. Dabei hat sich Wireshark für die Aufzeichnung und Auswertung von Netzwerkverkehr durchgesetzt. Wireshark bietet dabei auch ein Graphical User Interface (GUI) für die direkte Auswertung der Daten. Der grosse Nachteil von Wireshark ist aber die komplexe Handhabung. Das GUI bietet zwar viele Analysemöglichkeiten, die sind aber nicht immer intuitiv auffindbar. Mit Wireshark ist eine genaue Analyse erst möglich, sofern das benötigte Knowhow angeeignet wurde. Anderenfalls sind Auswertungen sehr zeitaufwändig.
Netzwerkverkehr kann für spätere Analysezwecke auch in Dumps, sogenannten Packet Captures, abgelegt werden. Hierzu existiert eine eigene Dateiendung Packet Caputre (PCAP) oder Packet Caputre Next Generation (PCAPNG).
Wireshark bietet zusätzlich zur GUI- noch eine CLI-Anwendung, namentlich Tshark. Tshark kann wie schon Wireshark Netzwerkverkehr aufzeichnen, gibt den aufgezeichneten

Traffic aber auf der Konsole aus.
Tshark ermöglicht auch die Ausgabe eines aufgezeichneten Capture auf der Konsole. Dabei können einzelne Attribute oder direkt ganze Packets ausgegeben werden.

3 Anforderungen

3.1 Personas

3.1.1 Persona 1

Persönliches Profil

Jürg gibt im Rahmen seiner Tätigkeit als Informatiker regelmässige Vorträge und Schulungen zum Thema Datenschutz.

Situation

Um den Teilnehmern aufzuzeigen, was bei alles bei der Nutzung eines Netzwerks übertragen wird greift er oft auf Wireshark zurück. Dies sieht er jedoch nicht als optimales Tool, da die dargestellten Informationen für die Teilnehmer oft nur schwer erkennbar sind.

Szenario

Im Rahmen einer Vorlesung soll es möglich sein den Netzwerkverkehr aufzuzeichnen und auswerten zu lassen. Zudem wäre es wünschenswert, wenn auch zu früheren Zeitpunkten aufgezeichneter Netzwerkverkehr ausgewertet werden könnte.
Für die Präsentation in Schulungen oder Vorlesungen werden diverse Informationen betreffend dem Netzwerkverkehr benötigt. So sollte es möglich sein zu sehen, wohin der meiste Verkehr geht beziehungsweise kommt, welche Secure Socket Layer (SSL)/Transport Layer Security (TLS)-Versionen und Cipher Suites verwendet wurden und um welche Art von Verkehr es sich handelt (z.B. Werbung oder Zugriff auf schädliche Websites). Wenn möglich sollen Informationen wie veraltete Cipher Suites oder Verbindungen auf Endpunkte, welche bekannterweise bösartig sind, farblich hervorgehoben werden, um die Schulungsteilnehmer diesbezüglich zu sensibilisieren.
Um aufzuzeigen, wo auf der Welt sich die entsprechenden Endpunkte befinden sollte es möglich sein, die ungefähre Location der Endpunkte anzuzeigen. Die Möglichkeit, nur Verkehr von einem spezifischen Endpunkt zu betrachten ist für einen solchen Einsatz essenziell. Zusätzliche Möglichkeiten zur Einschränkung der Auswertung würden die Einsatzbereiche noch erweitern.

3.1.2 Persona 2

Persönliches Profil

Patrick studiert Elektrotechnik. Sein Interesse für Technische Dinge hat dazu geführt, dass er Zuhause ein Netzwerk voller Internet of Things (IoT) Geräte besitzt, mit welchen er seine Wohnung automatisiert und durch eine Vielzahl an Sensoren überwacht.

Situation

In letzter Zeit hat er jedoch immer wieder davon gehört, dass einige Geräte auch Daten übermitteln, welche sie nicht übermitteln sollten, weshalb er sich nun sorgen über den Datenverkehr in seinem Heimnetzwerk macht.

Szenario

Zum Aufspüren von Unstimmigkeiten in einem Heimnetzwerk soll es möglich sein den Netzwerkverkehr über einen längeren Zeitraum zu analysieren. Um Erkenntnisse betreffen Geräten zu erhalten, welche sich auffällig verhalten, ist es wichtig die Anzahl Pakete und die Menge an Übermittelten Daten an verschiedene Endpunkte von einem spezifischen Gerät aus zu sehen.
Nebst diesen Informationen wäre es für den Heimnutzer wünschenswert, wenn potenziell gefährliche Endpunkte in der Darstellung farblich hervorgehoben werden.

3.1.3 Persona 3

Persönliches Profil

Petra arbeitet in einem Schweizer KMU in der IT-Abteilung und ist zuständig für die interne IT.

Situation

Seit einiger Zeit hat sie den Verdacht, dass Daten unerlaubt aus dem Firmennetzwerk abtransportiert werden. Ihre Anträge für den Kauf einer besseren Firewall, welche diese Problematik etwas abschwächen könnte, wurden bisher immer abgelehnt. Nun ist sie auf der Suche nach einer Möglichkeit, ihre Problematik zu visualisieren, um ihren Chef vom Kauf der neuen Hardware überzeugen zu können.

Szenario

Als Vorarbeit für die Analyse des Netzwerkverkehrs einer Firma wird über eine längere Zeitspanne der Netzwerkverkehr zwischen Router und Core-Switch aufgezeichnet. Die daraus resultierenden Daten werden danach ausgewertet.
Um den Chef von der Notwendigkeit der Anschaffung einer neuen Firewall zu überzeugen, benötigt sie eine visuelle Auflistung der übermittelten und empfangenen Daten ins Internet. Dabei möchte sie die jeweiligen Endpunkte im Internet filtern oder zusätzlich

sortieren können, sodass eine gezielte Auswertung ermöglicht werden kann. Allenfalls gefährliche Endpunkte sollen farblich hervorgehoben werden.
Weiter wäre eine zeitliche Filterung erwünscht, allenfalls mit vordefinierten Uhrzeiten, sodass über mehrere Tage, über gewisse Zeitspannen die Daten ausgewertet werden können.

3.1.4 Persona 4

Persönliches Profil

Pia arbeitet als IT Consultant und ist in vielen KMUs für die Verwaltung der Server verantwortlich.

Situation

Kürzlich fand sie bei einem Kunden heraus, dass der Webserver für die Lohnabrechnung noch immer veraltete Cipher Suites verwenden. Pia sucht nun nach einer einfachen Möglichkeit, Server der Kunden mit veralteten TLS/SSL-Version oder Cipher Suites zu erkennen.

Szenario

Durch aufgezeichneten Netzwerkverkehr soll eine schnelle Detektion von unsicheren Cipher Suites ermöglicht werden. Die Empfehlung, ob eine ausgehandelte Cipher Suite noch verwendet werden soll, muss dabei von einer öffentlichen Quelle stammen.
Mithilfe des Tools soll auch eine schnelle, wiederholte Überprüfung ermöglicht werden.

3.2 Non-Functional Requirements

Für die Klassifizierung der jeweiligen Non-Functional Requirements (NFRs) wird auf FURPS zurückgegriffen. FURPS steht für **F**unctionality, **U**sability, **R**eliability, **P**erformance, **S**upportability zusammensetzt. FURPS wurde von HP im Jahre 1992 entwickelt, um die Softwarequalität zu steigern. Dabei werden die einzelnen Requirements mit dem jeweiligen Buchstaben von FURPS verbunden.

#	FURPS	Titel	Beschreibung
1	**F**unctionality	Webzugriff	Das Frontend der Applikation soll per Webbrowser erreichbar sein.
2	**F**unctionality	HTTPS	Die Kommunikation zum Frontend soll per Hypertext Transfer Protocol Secure (HTTPS) stattfinden können.
8	**F**unctionality	Docker	Die Architektur soll so gestaltet sein, dass ein Docker Deployment möglich wäre.
3	**U**sability	Feedback Time	Der Benutzer soll nach spätestens fünf Sekunden ein Feedback zu jeglichen getätigten Interaktionen erhalten.
4	**U**sability	Access	Die Applikation soll vom Netzwerk aus erreichbar sein.
5	**R**eliability	24/7	Die Applikation soll für einen möglichen 24/7-Betrieb ausgelegt sein.
6	**P**erformance	Waiting Time	Das Wechseln zwischen jeglichen Seiten darf maximal eine Sekunde betragen.
7	**P**erformance	Filter Time	Die Dauer bis zur Anzeige von grafischen Statistiken darf nach der Filterung maximal 5 Sekunden betragen.
9	**S**upportability	Open Source	Die Applikation soll so geschrieben sein, dass für das Open Source Projekt weitere Funktionalitäten ermöglicht werden können.

Tabelle 3.1: Non-Functual Requirements

4 Lösungsvorschlag

4.1 Screens

Anhand der definierten Aufgabenstellung wurden zu Beginn der Arbeit einzelne Screens gezeichnet, welche die Interaktionen zur umzusetzenden App aufzeigen sollten. Mithilfe der Screens wurde die Richtung, in welche das User Interface (UI) gehen sollte, definiert. Der daraus resultierende Prototyp wurde in den Semesterwochen 2 besprochen, auf die Semesterwoche 3 überarbeitet und danach als gut befunden.

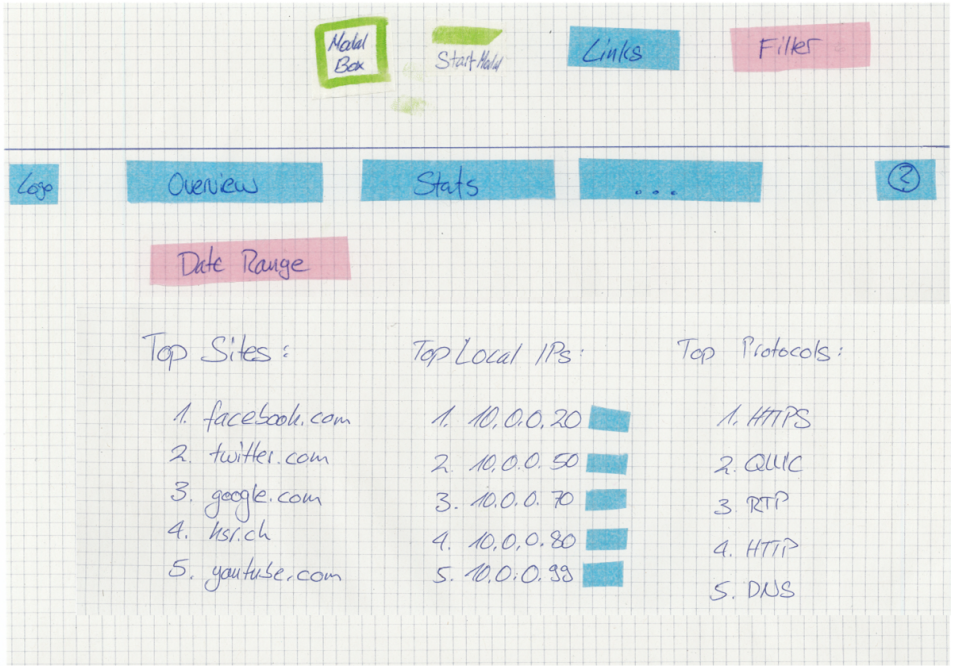


Abbildung 4.1: Erstellter Screen für die Übersicht des Netzwerkverkehrs

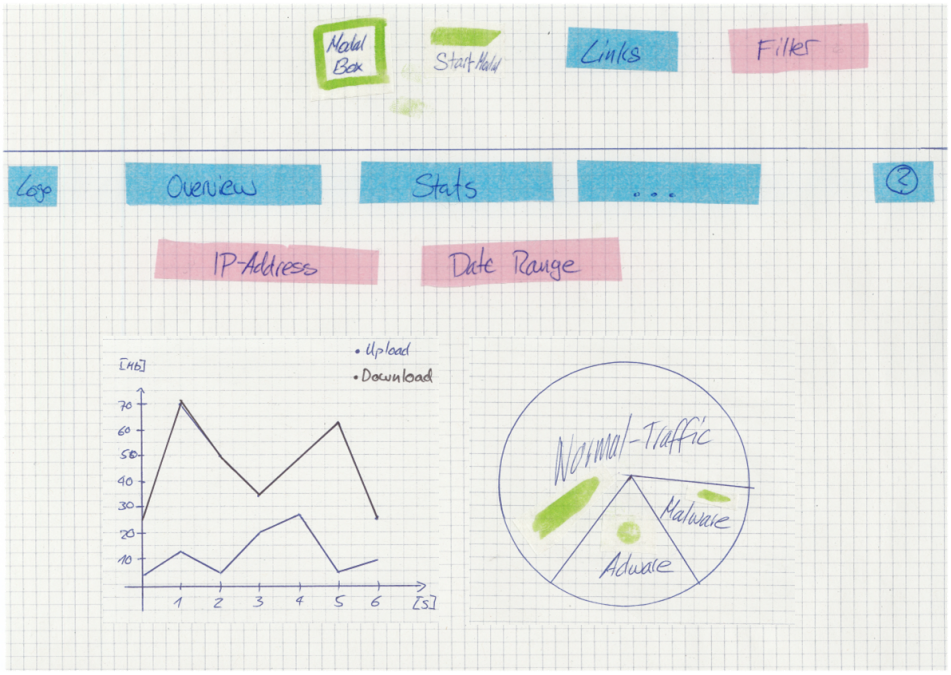


Abbildung 4.2: Erstellter Screen für die Ansicht der Statistiken des Netzwerkverkehrs

Alle Screens sind im Anhang in einer klickbaren Excel-Datei zu finden.

4.2 Splunk

Zu Beginn der Arbeit war die Idee, selbst eine vollständige Applikation mit eigenem Frontend, Backend und einer angebundenen Datenbank zu entwickeln. In der Semesterwoche 4 wurde diese Idee jedoch zu Gunsten eines anderen Lösungsansatzes, basierend auf einer Entwicklung einer Erweiterung zu einer bestehenden Software, verworfen.
Durch das von den Betreuern schon vorhandene Knowhow mit der Analyse Software Splunk, wurde entschieden, eine Erweiterung für Splunk in Form einer Splunk App zu programmieren. Splunk bietet dabei ein Grundgerüst mit Frontend, Backend und Datenbank. Mit diesem Ansatz kann mehr Fokus auf die eigentlichen Funktionen der Applikation gelegt werden. Die Screens konnten grösstenteils übernommen werden, da Splunk die definierten Darstellungsvarianten unterstützt.

4.3 Zuordnung von Netzwerkverkehr an Apps

Es wurde basierend auf Recherchen entschieden, dass das Feature zur Zuordnung von Netzwerkverkehr auf spezifische Applikationen nicht umgesetzt wird. Arbeiten an der Universität von Michigan [23] und der Cornell Universität [24] zeigen, dass für solche Funktionen ein grosser Aufwand nötig ist und dass in vielen Fällen keine eindeutige Zuordnung möglich ist.

4.4 Open Source

Die zu entwickelnde Applikation soll einer Open Source Lizenz unterstehen, sodass interessierte Anwender die Applikation ohne Einschränkungen nutzen und/oder weiterentwickeln können. Dies wurde in der Semesterwoche 2 festgelegt.

4.5 Anonymisieren der Daten

Das Anonymisieren der Daten wird bewusst nicht vorgenommen. Dies wurde in der Semesterwoche 5 und 6 so definiert. Wird die zu entwickelnde Applikation eingesetzt, hat die dafür zuständige Person sicherzustellen, dass alle durch die Analyse betroffenen Personen ihr Einverständnis gegeben haben oder das nötige Massnahmen zur Anonymisierung getroffen werden.
Das Auswerten von Daten ohne explizite, freiwillige Zustimmung der Mitarbeiter ist nach dem Schweizer Bundesgesetz in den meisten Branchen nicht zulässig [25] [26, S. 5].

5 Traffic Analyzer

Im Rahmen dieser Bachelorarbeit wurde **Traffic Analyzer**, ein Tool zum Auswerten von Netzwerkmitschnitten (.pcap-Dateien), entwickelt. Mithilfe von Traffic Analyzer soll es sowohl dem interessierten Heim- wie auch dem erfahrenen Netzwerk-Administrator möglich sein, sich schnell einen Überblick über den Netzwerkverkehr zu verschaffen.

5.1 Aufbau

Traffic Analyzer ist als Splunk-App entwickelt worden. Das Frontend wurde in Splunks Simple Extensible Markup Language (XML) realisiert. Die Abfragen auf die Datenbank von Splunk werden über die von Splunk entwickelte Abfragesprache SPL definiert. Im Backend kommen Python und Bash-Skripts zum Einsatz.

5.2 Benutzung

Da Traffic Analyzer auf Splunk aufbaut, ist für die Benutzung eine Splunk-Installation erforderlich. Ist Splunk auf einem Linux-System bereits vorhanden, kann die App mittels der Datei traffic-analyzer.tar.gz, welche im Git-Repository von Traffic Analyzer zu finden ist, installiert werden. Wenn diese nicht der Fall ist oder auf eine Installation von Splunk verzichtet werden soll, kann die App auch mit Splunk in einem Docker-Container genutzt werden. Eine Anleitung dazu ist direkt im Git-Repository (`https://github.com/anjo-hsr/Traffic-Analyzer`) zu finden.

5.3 Einstellungen

Beim ersten Starten der Applikation in Splunk wird der Benutzer aufgefordert, einige grundlegende Konfigurationen vorzunehmen (Abbildung 5.1), beispielweise Application Programming Interface (API)-Keys, der Pfad der PCAP-Ordner oder zusätzliche Domain Name System (DNS)-Server für die Auflösung von internen Adressen. Über den Menüpunkt «Settings» können die Einstellungen jederzeit aufgerufen und angepasst werden.

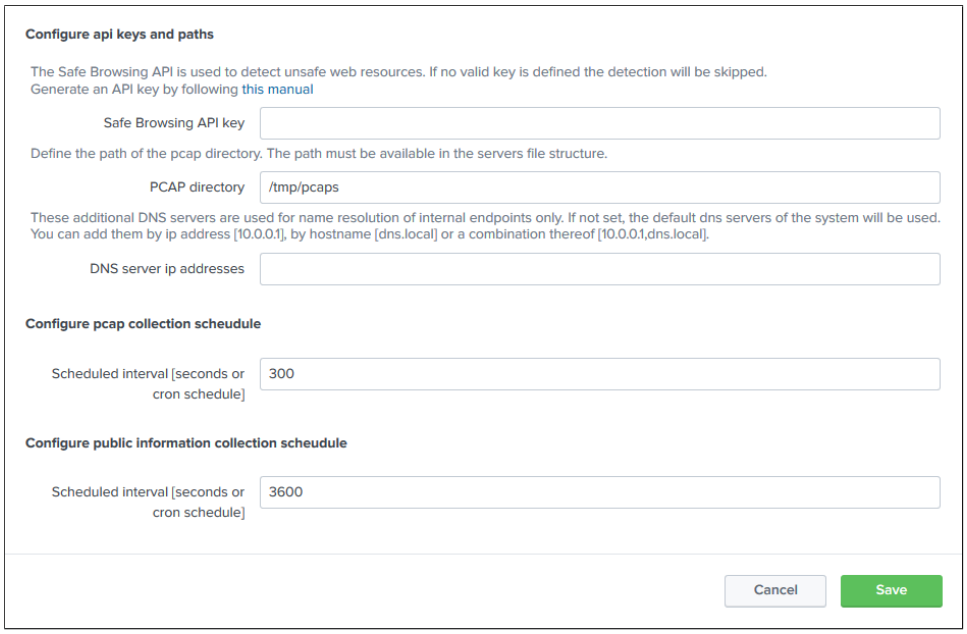


Abbildung 5.1: Settings: Traffic Analyzer

5.4 Filter

Die in Traffic Analyzer dargestellten Informationen können über Filter (Abbildung 5.2) eingeschränkt und für die eigenen Bedürfnisse angepasst werden.

Abbildung 5.2: Verfügbare Filter in Traffic Analyzer

5.4.1 Capture File

Der Filter für die Capture Files wird als Mehrfachauswahl implementiert. So hat der Benutzer die Möglichkeit, einen einzelnen Capture oder eine beliebige Kombination von verfügbaren Captures auszuwählen. Die Auswahl der verfügbaren Captures in der Mehrfachauswahl wird beim Laden eines Dashboards basierend auf den aktuell in Splunk eingelesenen Capture Files generiert.

5.4.2 MAC-Adresse

Der Filter für die Media Access Control (MAC)-Adresse (Abbildung 5.3) erlaubt es dem Benutzer, die Auswertungen auf einen spezifischen Endpunkt einzuschränken.

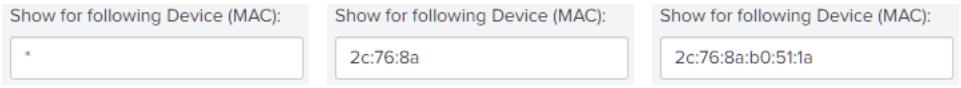


Abbildung 5.3: Verschiedene Möglichkeiten zur Filterung nach MAC-Adressen

So ist es zum Beispiel einfacher zu sehen, welche Cipher Suites und SSL/TLS-Versionen auf einem spezifischen Gerät zum Einsatz kommen oder an welchen Ort auf der Welt Verbindungen aufgebaut wurden. Dabei kann sowohl nach kompletten MAC-Adressen als auch nach Teilen davon, gefiltert werden.

5.4.3 Interne IP-Ranges

Die Filter für interne Internet Protocol (IP)-Ranges werden als Dropdown Menü implementiert und umfassen je zwei Auswahlmöglichkeiten (Abbildung 5.4).

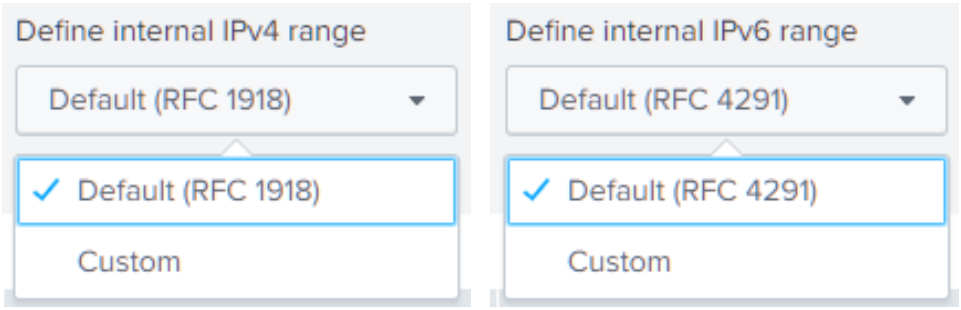


Abbildung 5.4: Optionen für Auswahl interner IP-Adressbereiche

Die Option «Default (RFC 1918)» für einen zusätzlichen IP-Range für Internet Protocol Version 4 (IPv4) behandelt alle in Requests For Comment (RFC) 1918 [27] definierten privaten IP-Adressbereiche als interne IP-Ranges. Wird die Option «Custom» ausgewählt kann zusätzlich noch ein weiterer IP-Range in Classless Inter-Domain Routing (CIDR)-Notation [28] angegeben werden (Abbildung 5.5).

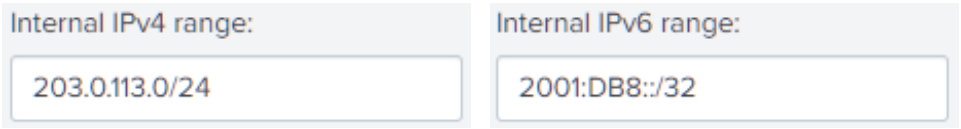


Abbildung 5.5: Eingabe eines zusätzlichen internen IP-Adressbereichs

Dasselbe gilt auch für die Auswahl eines zusätzlichen internen Internet Protocol Version 6 (IPv6)-Range, wobei da standardmässig die IPs als intern betrachtet werden, welche in RFC 4291 [29] als lokale IPs gelistet sind. Dies ist besonders von Relevanz, wenn öffentliche IP-Adressen im Local Area Network (LAN) verwendet werden.

5.4.4 Date & Time Picker

Alle Events können anhand des Date & Time Pickers gefiltert werden. Dabei ist der Standardwert «All time». Splunk bietet für diesen Filter schon vordefinierte Zeitbereiche an, mit welchen schnell der gewünschte Zeitraum angegeben werden kann.

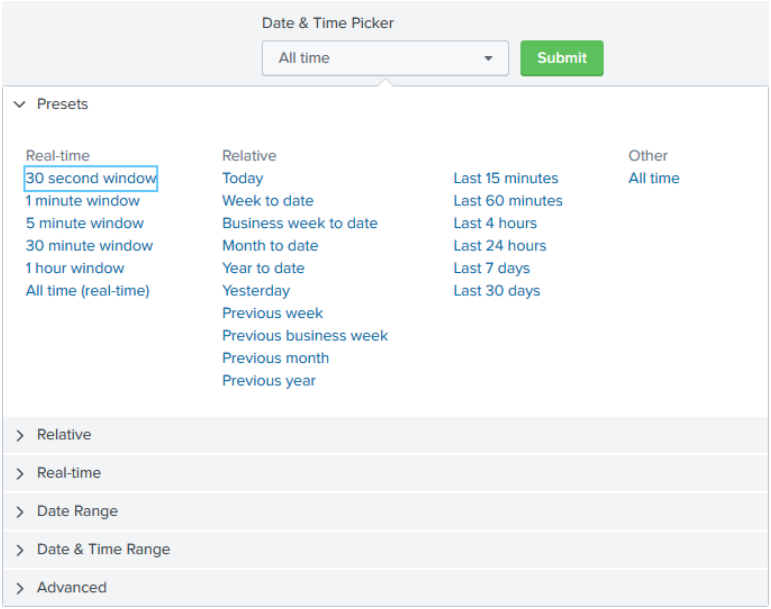


Abbildung 5.6: Filter für bestimmte Events anhand der Eventzeit

5.5 Gliederung

Die Auswertungen sind in verschiedene Teilgebiete aufgeteilt. Eine Übersicht soll zudem einen groben Überblick über einen Grossteil der Auswertungen bieten.

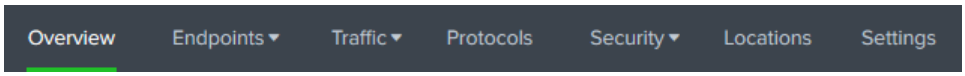


Abbildung 5.7: Navigation

5.5.1 Overview

Das Dashboard «Overview» bildet eine erste Anlaufstelle beim Auswerten der gesammelten Informationen. Dem Benutzer wird eine Übersicht über Informationen aus den Teilbereichen Endpoints, Protocols und Security präsentiert.

5.5.2 Endpoints

In der Kategorie «Endpoints» finden sich jene Dashboards, welche interne und externe Endpoints auflisten und weiterführende Informationen dazu darstellen.

Server Types

Das Dashboard «Server Types» zeigt in tabellarischer Form, welche Server eine bestimmte Aufgabe erfüllen. Der Benutzer hat über ein Dropdown die Möglichkeit auszuwählen, ob alle Server und ihre jeweiligen Funktionen angezeigt werden sollen oder nur jene welche eine durch ihn ausgewählte Funktion innehaben.

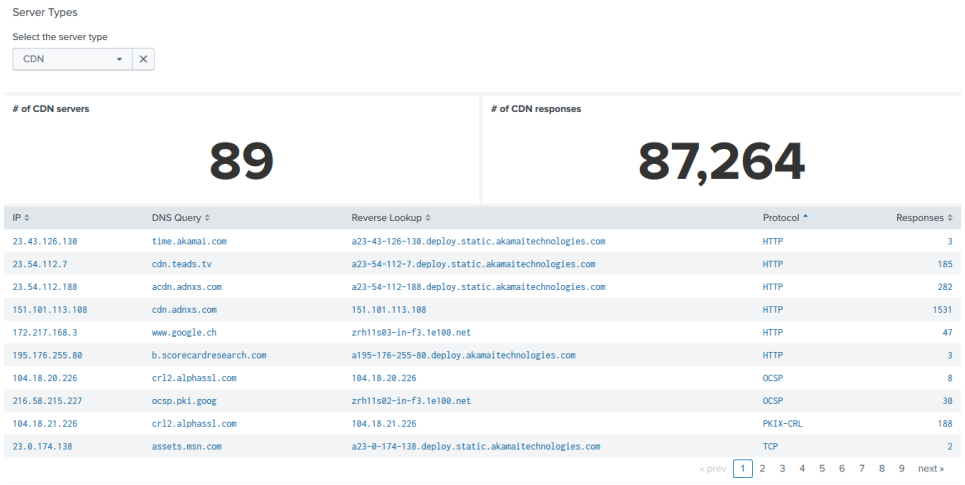


Abbildung 5.8: Dashboard: Server Types

Internal Endpoints

Das Dashboard «Internal Endpoints» zeigt alle vorkommenden Endpoints an, welche sich mit privaten IP-Adressen nach RFC 1918 oder RFC 4291, wie auch IP-Adressen, die sich in einem vom Benutzer spezifizierten, internen IP-Range befinden, melden. Zusätzlich zu den IPs werden dem User die MAC-Adresse und die Anzahl an ausgehenden und einkommenden Bytes pro Endpoint angezeigt.

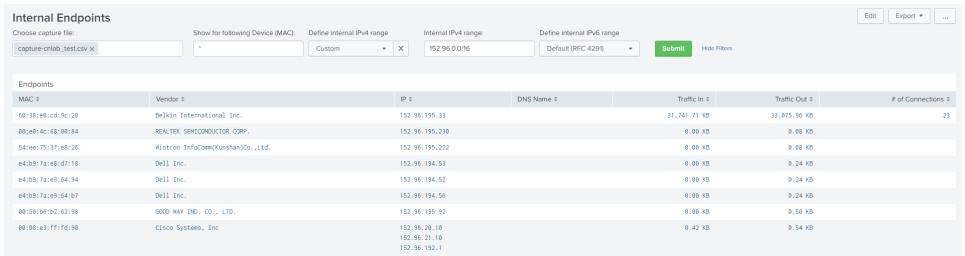


Abbildung 5.9: Dashboard: Internal Endpoints

External Enpoints

Das Dashboard «External Endpoints» zeigt alle vorkommende Endpoints im Internet an. Die zugehörige IP wird mit Informationen zum DNS-Namen und der Anzahl an ausge-

henden und einkommenden Byte ergänzt.

Abbildung 5.10: Dashboard: External Endpoints

5.5.3 Traffic

In der Kategorie «Traffic» sind jene Dashboards zu finden, welche den Traffic geordnet nach Kategorien wie Threat, Ad und normalem Traffic darstellen.

5.5.4 Stream Information

Im Dashboard «Stream Information» werden Informationen zu einzelnen Streams aufgezeigt. Dabei werden die übermittelte Datenmenge und die Dauer des Streams zwischen zwei Endpunkten in Form einer Tabelle dargestellt.

Traffic Types

Das Dashboard «Traffic Types» stellt die aufgezeichneten Verbindungen geordnet nach normalem Traffic, Ad Traffic und Malicious Traffic als Kuchendiagramm dar. Abhängig von der aus der IP-Adresse ermittelten Geolocation, werden diese Informationen nach geografischem Standort aufgeteilt und am entsprechenden Ort auf der Karte angezeigt.

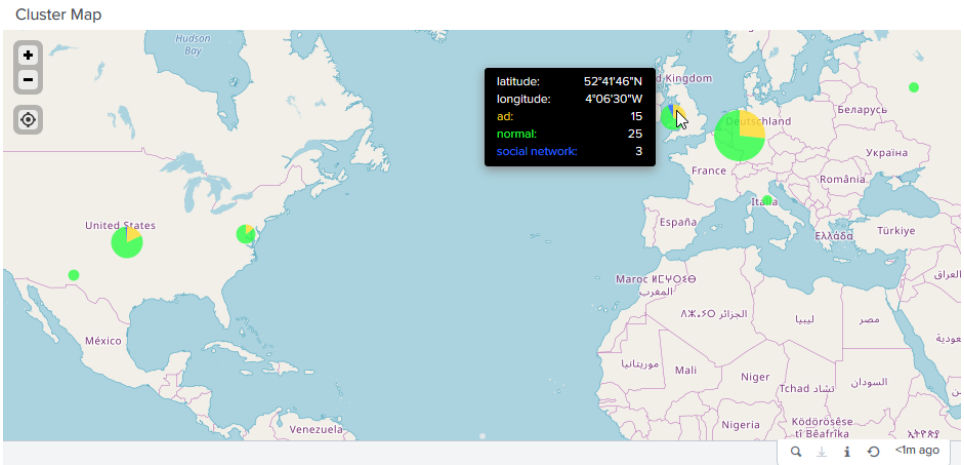


Abbildung 5.11: Kuchendiagramme für Regionen mit angesprochenen Endpunkten

Durch einen Klick auf ein Kuchendiagramm wird eine genaue Liste der Verbindungsziele in der angewählten Umgebung angezeigt.

Destinations between lat(45.00000, 67.50000) and long(-45.00000, 0.00000)

DNS Query	Reverse Lookup	Traffic Type
cms.quantserve.com	91.228.74.205	ad
crl3.digicert.com	93.184.220.29	normal
ad.360yield.com	ec2-52-48-121-18.eu-west-1.compute.amazonaws.com	ad
swisscom.demdex.net	ec2-52-211-104-45.eu-west-1.compute.amazonaws.com	ad
	40.112.75.175	normal
	pr-bh.pbp.vip.ir2.yahoo.com	normal
	xx-fbcdn-shv-01-frt3.fbcdn.net	normal
match.prod.bidr.io	ec2-52-212-115-169.eu-west-1.compute.amazonaws.com	ad
ad.turn.com	46.228.164.11	ad
	ec2-34-248-238-74.eu-west-1.compute.amazonaws.com	normal

« prev 1 2 3 4 5 next »

Abbildung 5.12: Endpunkte eines ausgewählten Kuchendiagramms

Ad Locations

Das Dashboard «Ad Locations» zeigt die verschiedenen Arten von Werbe- und Tracking-Verkehr verteilt auf die geografischen Standorte an.

Social Network Locations

Das Dashboard «Social Network Locations» zeigt die Anzahl an Server von sozialen Netzwerken in Ländern an.

Threat Locations

Das Dashboard «Threat Locations» zeigt die verschiedenen Arten von bösartigem Verkehr verteilt auf die geografischen Standorte an.

5.5.5 Protocols

Das Dashboard «Protocols» bietet eine Übersicht über die aufgezeichneten Protokolle nach Anzahl Pakete und Verkehrsvolumen als Kuchendiagramm. Zusätzlich wird auf einem Zeitstrahl dargestellt, wie die Nutzung der Protokolle sich innerhalb der Aufzeichnung über Zeit verändert hat.

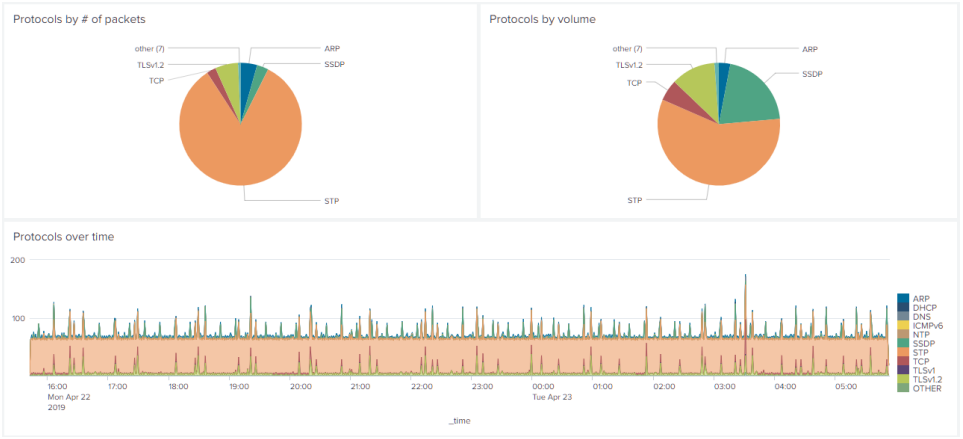


Abbildung 5.13: Dashboard: Protocols

5.5.6 Security

IPv6 Security

Das Dashboard «IPv6 Security» zeigt an, ob einzelne EUI-64 IPv6-Adressen im Netzwerk existieren, welche anhand der MAC-Adresse ermittelt wurden.

Abbildung 5.14: Dashboard: IPv6 Security

TLS Security

Das Dashboard «TLS Security» zeigt Informationen über das Verhältnis zwischen sicherer und unsicherer Kommunikation, verwendete Cipher Suites, TLS/SSL-Versionen und die Anzahl an unsicheren Cipher Suites an.
Die Kachel, welche die gesamte Anzahl an analysierten TCP Streams anzeigt (Abbildung 5.15), dient dazu, die restlichen Werte auf dem Dashboard besser einordnen zu können.

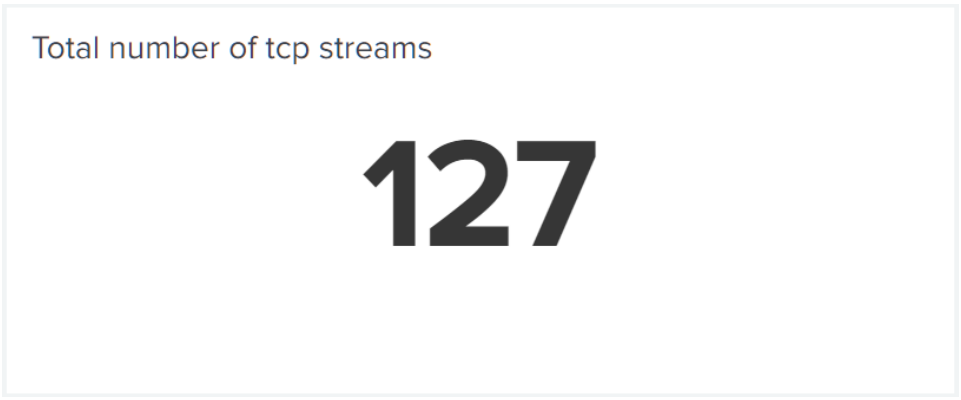


Abbildung 5.15: Gesamtanzahl an TCP Streams

Die Kachel mit den nicht zu empfehlenden Cipher Suites (Abbildung 5.16) zeigt als absoluten und als prozentualen Wert an, wie viele der gefundenen Cipher Suites im analysierten Capture von der Organisation Internet Assigned Numbers Authority (IANA) [30] nicht mehr zur Verwendung empfohlen werden.

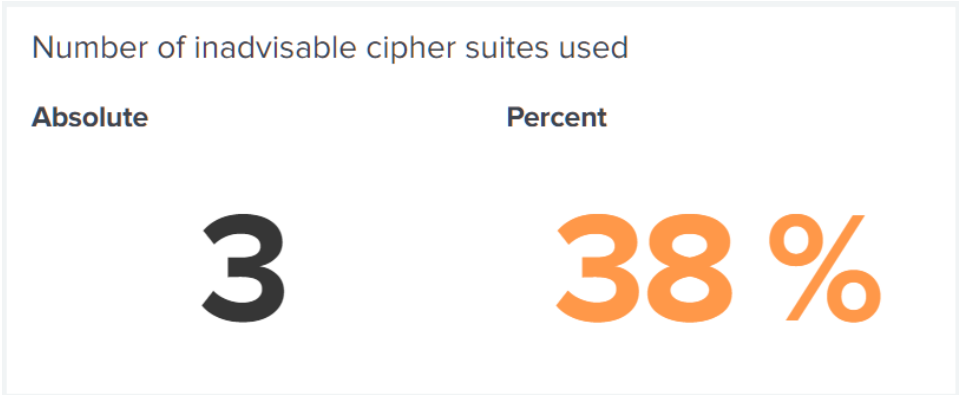


Abbildung 5.16: Nicht zu empfehlende Cipher Suites

Die Kachel zur Menge an unsicherem Verkehr (Abbildung 5.17) zeigt an, wie viele Pakete ungesichert übermittelt wurden und welchem Prozentsatz von der Gesamtzahl an übermittelten Paketen dies entspricht.

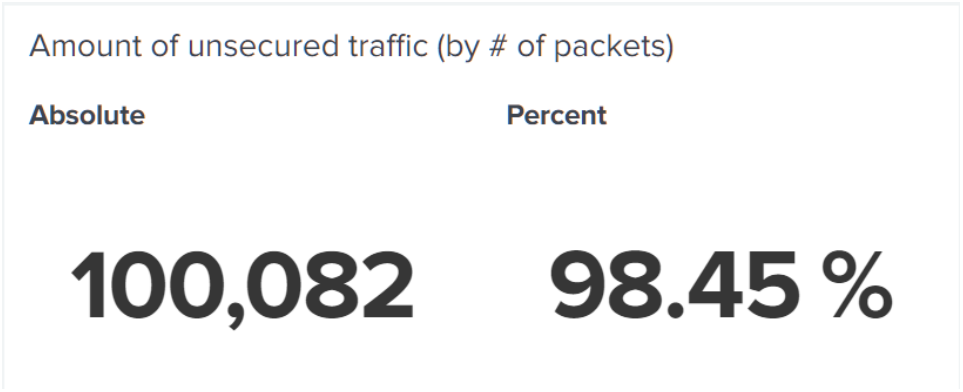


Abbildung 5.17: Unsicherer Verkehr nach Anzahl Paketen

In der Tabelle mit den verwendeten Cipher Suites (Abbildung 5.18) werden alle erkannten Cipher Suites gelistet. Zusätzlich wird angezeigt, für wie viele Verbindungen die jeweilige Cipher Suite verwendet wurde und ob die Cipher Suite noch zur Verwendung empfohlen wird. Cipher Suites welche laut IANA nicht mehr verwendet werden sollten, werden in der Spalte «Recommended» Rot markiert.

Commonly used cipher suites

Nr.	Description	Recommended	# of streams
49199	TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	Y	28
49195	TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	Y	10
49200	TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384	Y	7
4865	TLS_AES_128_GCM_SHA256	Y	4
157	TLS_RSA_WITH_AES_256_GCM_SHA384	N	3
47	TLS_RSA_WITH_AES_128_CBC_SHA	N	1
52392	TLS_ECDHE_RSA_WITH_CHACHA20_POLY1305_SHA256	Y	1
53	TLS_RSA_WITH_AES_256_CBC_SHA	N	1

Abbildung 5.18: Verwendete Cipher Suites

Die Tabelle mit den verwendeten TLS/SSL-Versionen (Abbildung 5.19) zeigt an, in welcher Anzahl von Verbindungen die jeweilige Version genutzt wird.

Commonly used SSL/TLS version

Version	Description	# of streams
0x00000303	TLS 1.2	51
0x00000304	TLS 1.3	4

Abbildung 5.19: Verwendete TLS/SSL-Versionen

Durch die Auswahl einer spezifischen Cipher Suite aus der Liste mit erkannten Cipher

Suites kann zusätzlich angezeigt werden, welche externen Endpunkte diese Cipher Suite verwenden und mit welchen internen Endpunkten so kommuniziert wurde (Abbildung 5.20).

Target IP	DNS Name	Nr.	Description	Recommended	Source IP
37.252.172.250	538.bm-nginx-loadbalancer.mgmt.fra1.adnexus.net	49195	TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	Y	152.96.195.33
83.150.0.51	speedtest.iway.ch	49195	TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	Y	152.96.195.33
216.58.215.226	zrh11s02-in-f2.1e100.net	49195	TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	Y	152.96.195.33
172.217.168.2	zrh11s03-in-f2.1e100.net	49195	TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	Y	152.96.195.33

Abbildung 5.20: Cipher Suite Drilldown

5.5.7 Location

Das Dashboard «Location» zeigt, welche Anzahl Endpunkte in welchen Ländern angesprochen wurden.

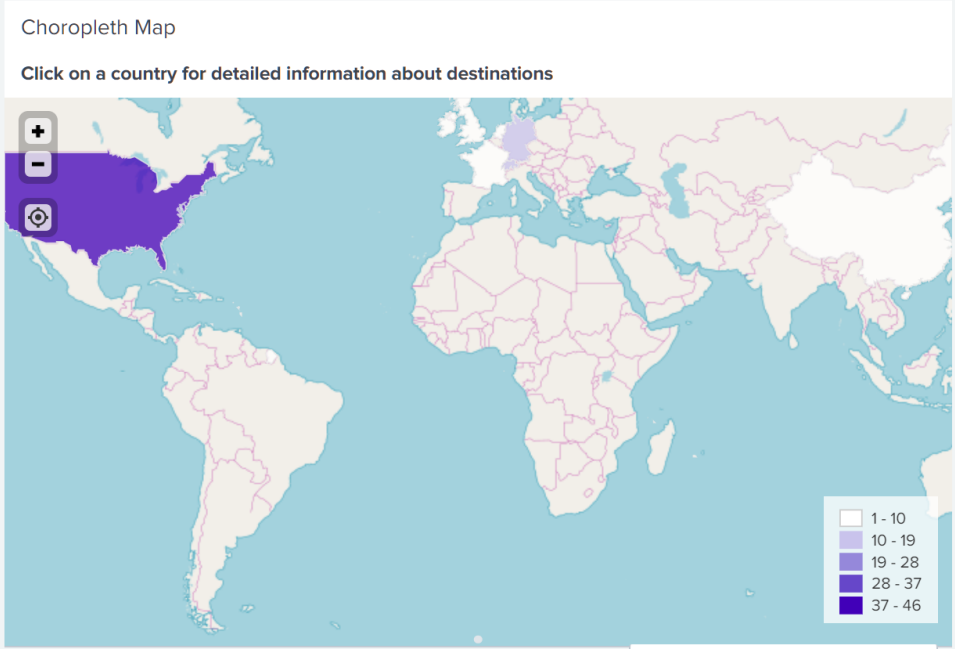


Abbildung 5.21: Länder mit angesprochenen Endpunkten

Durch einen Klick auf ein Land kann zudem eine Liste mit allen Verbindungszielen in dem Land und der Anzahl dahin aufgebauter Verbindungen angezeigt werden. Dabei wird auch ein Hinweis zu den vom EDÖB veröffentlichten Informationen zu den Datenschutzbestimmungen des jeweiligen Landes gezeigt. [31]

Destinations in Ireland

The Swiss government considers this country to have appropriate safety for personal data of natural persons.
Source: List of countries by the Swiss Federal Data Protection and Information Commissioner

DNS Query	Reverse Lookup	# of connections
logs1407.xiti.com	ec2-52-49-204-15.eu-west-1.compute.amazonaws.com	7
edge-star-mini-shv-01-amt2.facebook.com	edge-star-mini-shv-01-amt2.facebook.com	5
logs1407.xiti.com	ec2-54-171-180-56.eu-west-1.compute.amazonaws.com	2
logs1407.xiti.com	ec2-54-76-54-11.eu-west-1.compute.amazonaws.com	2
graph.facebook.com	edge-star-shv-01-amt2.facebook.com	2
api-prod.axonvibelabs.com	ec2-34-246-67-219.eu-west-1.compute.amazonaws.com	1
	ec2-52-49-85-36.eu-west-1.compute.amazonaws.com	1
zdbb.net	ec2-54-154-81-193.eu-west-1.compute.amazonaws.com	1

Abbildung 5.22: Endpunkte in einem ausgewählten Land

6 Begriffserklärungen

Für die Umsetzung der Splunk App werden zu bestehenden Captures weitere Informationen herbeigezogen. Diese Begriffe werden innerhalb dieses Kapitels behandelt.

6.1 Server Funktionen

6.1.1 DHCP Server

Dynamic Host Configuration Protocol (DHCP)-Server sind innerhalb eines Netzwerkes für die Vergabe von IP-Adressen verantwortlich. Sie händigen Clients, welche keine statische IP-Adresse hinterlegt haben, für einen gewissen Zeitrahmen eine IP-Adresse aus. Diese IP-Adresse dient später zur Kommunikation zwischen IP-Subnetzen.
Der Ablauf einer normalen DHCP-Abfrage durchläuft dabei folgende Schritte. [32]

- Client
  - **DHCP Discover**: Der Client initiiert den DHCP-Prozess, indem er ein DHCP Discover ins Netzwerk sendet.
- Server
  - **DHCP Offer**: Alle DHCP-Server, welche im Netzwerk den DHCP Discover erhalten habe, beantworten diese mit einem DHCP Offer, in welcher die vom Server vorgeschlagene IP-Adresse beinhaltet ist. Dabei teilt er dem Client gleich seine eigene IP-Adresse mit.
    Zusätzlich sendet der DHCP-Server weitere Informationen, sogenannte Options, im Packet mit. Diese Options werden dabei für die weitere Kommunikation ausserhalb des verbundenen Netzwerkes benötigt [1] [2] [3]. Andere Options können dabei für die Konfiguration von Diensten auf dem jeweiligen DHCP-Client genutzt werden [4] [5] . [33]
- Client

[1] Option 1: Subnetzmaske
[2] Option 3: Default Gateway
[3] Option 5: Name Server
[4] Option 69: Simple Mail Transfer Protocol (SMTP) Server
[5] Option 42: Network Time Protocol (NTP) Server

  - **DHCP Request**: Nach dem Empfangen eines DHCP Offers, testet der Client, ob antwortet der Client allen anderen Clients mit einem DHCP Request, mit welchem er ihnen mitteilt, dass er eine IP-Adresse erhalten hat. Zusätzlich stehen im Packet Informationen zum gewählten DHCP-Server.

- Server
  - **DHCP ACK**: Der gewählte DHCP-Server sendet dem Client nochmals eine Zusammenfassung mit den ausgehändigten Informationen zu.

- Client
  - **Gratuitous ARP**: Nach dem Erhalt des DHCP ACK sendet der neue Client im Netzwerk die erhaltene IP-Adresse mittels Address Resolution Protocol (ARP) ins Netzwerk. Dadurch könnte sich ein anderes System, welches dieselbe IP-Adresse nutzt, melden, um eine spätere IP-Kollision zu vermeiden. Der neue Client verwirft danach die IP-Adresse, um mit einem DHCP decline eine neue IP-Adresse anzufragen und dem DHCP-Server mitzuteilen, dass die IP-Adresse bei einem anderen System verwendet wird.

Das Problem am DHCP-Protokoll ist, dass der Client nur das erste DHCP Offer weiterverarbeitet und weitere DHCP Offers verwirft. Durch einen DHCP-Spoofing-Angriff, bei welchem ein DHCP-Server schneller mit einer DHCP Offer antwortet, können dem Client falsche Informationen mitgeteilt werden. Dadurch wäre es dem Angreifer möglich Netzwerkverkehr über einen falschen Router zu leiten oder einen DNS-Server des Angreifers zu hinterlegen. Die falschen Systeme können danach für weitere Man-In-The-Middle-Attacken verwendet werden. [34] Ein DHCP-Poisoning-Angriff kann auf konfigurierbaren Switches mit der DHCP Snooping Option unterbunden werden. [34]

6.1.2 DNS Server

DNS-Server sind für die Auflösung von Domainnamen und IP-Adressen zuständig. Bei einer einem Forward Lookup fragt der Client anhand eines Domainnamens eine IP-Adresse an. Bei einem Reverse Lookup anhand einer IP-Adresse einen Domainnamen. Eine Forward Lookup DNS-Abfrage läuft wie folgt ab:

- Client
  - **DNS standard query**: Der Client sendet dem DNS Server ein DNS standard query. Standard queries werden nur für Forward Lookups, Auflösung von Domainnamen in IP-Adresse, verwendet [35]. Im Packet definiert der Client auch den gewünschten Reponse Type [36].

- Server
  - **DNS standard query response**: Der DNS-Server beantwortet das Standard query mit einem standard query response. Eine erfolgreiche Auslösung returniert die aufgelöste IP-Adresse, wie auch den zuvor angefragten

Domainnamen. Die Antwort kann aber auch mehrere IP-Adressen beinhalten.
Eine nicht erfolgreiche Auflösung sendet neben der gewünschten Anfrage auch den jeweiligen Fehler im Response Code zurück [35].

Das DNS-Protokoll ist, wie schon bei DHCP, mit einem DNS-Spoofing-Angriff verwundbar.
Jede DNS-Anfrage besitzt eine Transaction ID, welche einen DNS Request mit einer DNS Response verbindet. Sofern es einem Angreifer gelingt, vor dem DNS-Server eine Antwort mit richtiger Transaction ID dem Client zu übermitteln, wird dieses Packet vom Client als korrekte Antwort angesehen. So kann der Angreifer die spätere Anfrage auf ein eigenes System weiterleiten.
Durch die zufällige Vergabe von DNS Transaction IDs kann das Erraten der ID verhindert werden, sofern eine gute Random Quelle genutzt wird [37]. Auch die Nutzung von DNS over TLS (DoT) oder DNS over HTTPS (DoH) kann einen DNS-Spoofing-Angriff verhindert [38] [39]. Durch die verschlüsselte Übertragung der DNS-Pakete, kann der Angreifer keine eigenen Packets mehr ins Netzwerk senden, welche von Client angenommen werden.
Der DNS Cache Poisoning Angriff ist dabei das Setzen von falschen Informationen im Cache eines DNS Server. Wie schon Clients speichert ein DNS-Server DNS-Anfragen im Cache für die Time to live (TTL) ab. Mithilfe von Domain Name System Security Extensions (DNSSEC) kann DNS Cache Poisoning verhindert werden, da DNS-Antworten signiert returniert werden [40].

6.2 Server Typen

6.2.1 CDN Server

Content Delivery Network (CDN)-Server sind Server Systeme, welche über mehrere Regionen verteilt sind. Sie liefern überall den identischen Content an Clients aus. Dadurch wird eine kleinere Round Trip Time (RTT) bis zum Server benötigt, was die User Experience (UX) verbessert. Aktuell wird etwa die Hälfte des weltweite Netzwerkverkehrs mithilfe von CDNs-Servern übermittelt [41].
CDN-Server können für unterschiedliche Zwecke eingesetzt werden. Ob Streaming oder anderer Webcontent, solange der Inhalt der abzurufenden Daten über mehrere Regionen gleich sein sollte, macht es Sinn einen CDN-Dienst zu betreiben.

6.3 Stream ID

Stream Identifiers (IDs) werden anhand der Verbindung zwischen zwei Systemen ermittelt. Streams IDs werden sowohl für Transmission Control Protocol (TCP)-, als auch für User Datagram Protocol (UDP)-Verbindungen erstellt. Unterschiedliche Streams werden anhand der Source und Destination Sockets erkannt. Ein Socket wird durch die Kombination einer IP-Adresse mit einem zugehörigen Port definiert.

Stream IDs werden aufsteigend nummeriert und beginnen bei null. Stream IDs können zwar von Netzwerksniffern festgelegt werden, sind aber auf die aktuell eingelesene Datei oder dem aktuell laufendem Caputure beschränkt.

6.4 Geo Standort

Anhand einer öffentlichen IP-Adresse lässt sich der ungefähre Standort eines Systems bestimmen. Mithilfe der Standortinformationen ist es möglich Werbung spezifisch auf einen Standort zu zeigen oder Antworten von Suchbegriffen in einem Radius zu beschränken.
Den Standort eines Devices kann bei Anbietern über Webseiten oder APIs, abgefragt werden. Wie die Anbieter an die Informationen gelangen, wird nicht kommuniziert. Kostenlose Abfragen sind jedoch an ein Limit von Anfragen pro Zeiteinheit gebunden [6] [7] [8].
Bei Übertretungen der Limite werden weitere Anfragen abgeblockt [42]. Sofern mehr Anfragen in der festgelegten Zeitspanne benötigt werden, offerieren viele Anbieter einzelne Preismodelle, mit welchem höhere Limiten erkauft werden können [9] [10].
Als Alternative zu den einzelnen Anfragen über eine API gibt es auch vereinzelte Anbieter, welche Datenbanken zum Download bereitstellen. Mit der Datenbank sind die Abfragen offline und ohne zeitliche Einschränkung möglich. Hierbei wird wieder zwischen frei verfügbaren Datenbanken und kostenpflichtigen Datenbanken unterschieden. Kostenlose Geo-IP-Datenbanken haben den Nachteil, dass sie in einem weit grösseren Zeitintervall aktualisiert werden. Kostenlose Datenbanken beinhalten in der Regel auch weit weniger Informationen als das kostenpflichtige Pendant. [45]
Die Qualität solcher Online-Quellen hängt stark von der Internet-Infrastruktur des Landes ab [46]. Die Lokalisierung eines Gerätes mit einer öffentlichen IP-Adresse auf ein Land ist weit weniger schwierig als beispielsweise auf eine bestimmte Region oder Ortschaft [47]. Das Problem ist, dass sich die IP-Adressen, welche von Internet Service Providers (ISPs) per DHCP verteilt werden, in kurzen Zeiträumen ändern können. Dadurch kann beispielsweise die IP-Adresse, welche zuvor einem in Zürich stationierten Gerätes zugeordnet war, neu für ein anderes Gerät in Winterthur vergeben werden.
Bei der Lokalisierung einzelner Systeme greifen die Anbieter auf Datenbanken zurück, welche auch Standard-Orte beinhalten können. Dadurch sind Fehler der Anbieter nicht auszuschliessen [48] [49].

6.4.1 Clientseitige Ortungslösung

Solche Fehler können nur umgangen werden, sofern der Client selbst weitere Informationen liefert. Dies kann mittels Global Positioning System (GPS)-Modul oder mit weite-

[6] KeyCDN: 3 Anfragen pro Sekunde [42]
[7] ipgeolocation: 1'000 Anfragen pro Tag [43]
[8] ipstack: 10'000 Anfragen pro Monat [44]
[9] ipgeolocation: 15.- $ pro Monat für 150'000 Anfragen pro Monat + 5.- $ für weitere 50'000 Requests [43]
[10] ipstack: 9.99 $ pro Monat für 50'000 Anfragen pro Monat [44]

ren Daten von umliegenden Netzwerken erreicht werden. Für einen genaueren Standort können Smartphones beispielsweise die aktuell verbundenen Sendeantenne des Mobilnetzwerkes, sowie weitere in der Nähe erreichbaren Sendeantennen dem Server übermitteln. Auch mithilfe von umliegenden Wireless Local Area Network (WLAN)-Service Set Identifiers (SSIDs) ist eine genauere Lokalisierung möglich. Solch eine Lokalisierung kann mit der JavaScript Geolocation oder der Google Geolocation API umgesetzt werden [50] [51].

6.4.2 Anycast Adressen

Anycast definiert die Nutzung einer IP-Adresse auf mehreren Systemen, welche die gleiche Funktion erfüllen. Bei einer Anfrage wird an eine Anycast Adresse wird, festgelegt durch Routing-Regeln auf den Routern, der am nächstgelegene Server antworten. So wird die RTT verkürzt. Das Routing an eine Anycast Adresse wird für öffentliche Adressen von den ISPs vorgenommen.
Für IPv4-Adressen existieren dabei keine fix zugeteilten Adressen. Eine Anycast Adresse kann dabei irgendeine IPv4-Adresse sein. Auch für IPv6 existiert kein definierter Anycast Addressrange [52], auch wenn dies in einem Proposed RFC Standard erwähnt wird [53].
Die Erkennung einer solchen Anycast Adresse somit nur durch Pings von mehreren Standorten aus möglich. Eine solche Applikation stellt KeyCDN mit ihrem Ping Test zur Verfügung [54]. Über die Ping Test Webseite ist es möglich eine definierte IP-Adresse oder einen Domainnamen von verschiedenen KeyCDN Standorten aus zu pingen.
Normale, öffentliche IP-Adressen sind jedoch an die geopraphischen Distanzen zwischen den Ländern gebunden. Dabei ist in etwa mit einer RTT von etwa 100ms zwischen Zürich und New York zu rechnen [55], zwischen Zürich und Frankfurt in etwa mit 6ms.

LOCATION	IP	REQUESTS	MIN	MAX	AVG
Frankfurt	62.2.156.89	3	17.417 ms	19.539 ms	18.236 ms
New York	62.2.156.89	3	124.368 ms	129.327 ms	127.233 ms
Miami	62.2.156.89	3	145.511 ms	158.826 ms	150.827 ms
Dallas	62.2.156.89	3	150.792 ms	155.319 ms	152.527 ms
San Francisco	62.2.156.89	3	186.570 ms	193.289 ms	190.292 ms
Seattle	62.2.156.89	3	203.223 ms	216.429 ms	209.723 ms
Toronto	62.2.156.89	3	143.531 ms	146.386 ms	145.153 ms
London	62.2.156.89	3	33.093 ms	35.219 ms	34.075 ms
Paris	62.2.156.89	3	29.417 ms	34.016 ms	31.998 ms

Abbildung 6.1: Bildausschnitt der Erkennung der normalen, öffentlichen Adresse 62.2.156.89 mithilfe des Ping Test von KeyCDN [54]

Anhand er zurückgegebenen, durchschnittlichen Ping Zeit, ist erkennbar, dass die Systeme hinter der Anycast Adresse 8.8.8.8 nicht nur in New York, sondern auch in weiteren Ländern vertreten sind.

LOCATION	IP	REQUESTS	MIN	MAX	AVG	STD DEV
Frankfurt	8.8.8.8	3	12.454 ms	12.472 ms	12.463 ms	0.091 ms
New York	8.8.8.8	3	1.600 ms	2.381 ms	1.958 ms	0.326 ms
Miami	8.8.8.8	3	0.818 ms	0.828 ms	0.824 ms	0.004 ms
Dallas	8.8.8.8	3	0.932 ms	1.442 ms	1.123 ms	0.229 ms
San Francisco	8.8.8.8	3	2.802 ms	3.075 ms	2.917 ms	0.115 ms
Seattle	8.8.8.8	3	0.422 ms	0.491 ms	0.455 ms	0.033 ms
Toronto	8.8.8.8	3	0.779 ms	1.264 ms	0.960 ms	0.216 ms
London	8.8.8.8	3	0.327 ms	0.683 ms	0.446 ms	0.167 ms
Paris	8.8.8.8	3	1.023 ms	1.048 ms	1.033 ms	0.028 ms

Abbildung 6.2: Bildausschnitt der Erkennung einer öffentlichen Anycast Adresse 8.8.8.8 mithilfe des Ping Test von KeyCDN [54]

Anycast Adressen müssen aber nicht über mehrere Länder verteilt sein. Eine Anycast Adresse kann auch von zwei Systemen innerhalb der Schweiz genutzt werden. Durch die weit kleinere Distanz zwischen den Server-Systemen, müsste auch die Distanz zwischen den Ping Servern reduziert werden, sodass ein anderes ISPs-Routing überhaupt angewendet werden kann, um Unterschiede der RTT zu ermitteln.
Dadurch ist eine exakte Lokalisierung einer Anycast Adresse nicht möglich.

6.5 Virtual Host - Webserver

Webserver sind heutzutage nicht mehr nur für eine einzige Webseite zuständig. Häufig werden mehrere Webseiten auf einem einzelnen Server betrieben. Dies ist bei Hosting-Providern meist Standard. Die Verwaltung von mehreren Webseiten wird mit sogenannten Virtual Hosts vorgenommen. Virtual Host sind Einträge in der Konfiguration der Webserver, welche gewisse IPs oder Domains an eine Webseite binden. Es können aber auch mehrere Domains oder IPs auf eine Webseite verlinken. Anhand der vom Client angefragten Webseite entscheidet der Hosting Server, welche Webseite er zurücksenden muss. [56]

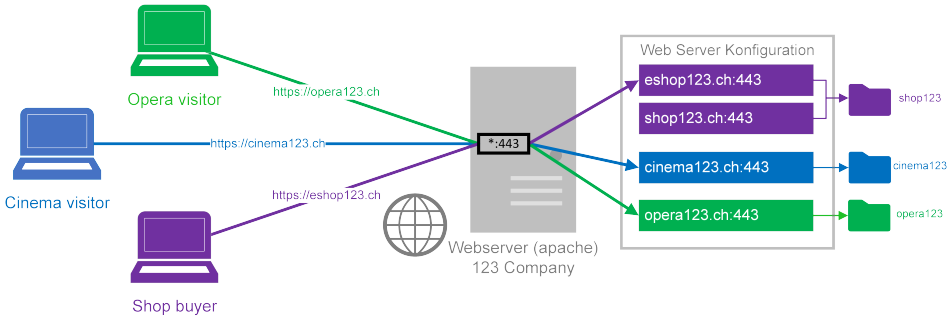


Abbildung 6.3: Vermittlung an die virtuellen Hosts basierend der angefragten Webseiten

Durch das gleichzeitige Hosten von mehreren Domains kann die benötigte Anzahl von öffentlichen IP-Adressen eingeschränkt werden. Bei einem Apache Server besteht eine Einschränkung von maximal 64 Virtual Hosts pro Hosting System. Diese Zahl kann in der Konfiguration jedoch erhöht werden [57].

6.6 Reverse Proxy

Ähnlich zur Nutzung von Virtual Hosts ist die Verwendung eines Reverse Proxys. Ein Reverse Proxy besitzt jedoch in seiner Webserver Konfiguration Einträge, welche Anfragen an andere Systeme weiterleitet. Ein Reverse Proxy ist dadurch nur der Mittelmann einer Anfrage, welche er nicht selbst beantwortet. Die Systeme, an welche er die Anfrage weiterreicht, sind meist in einem, nicht von aussen erreichbarem Netzwerk lokalisiert. Ein Reverse Proxy benötigt durch das Weiterreichen von Requests weniger Hardware. Zudem ist eine plattformunabhängige Entwicklung hinter einem Reverse Proxy möglich, da diese Systeme nicht die gleiche Webserver-Version bieten müssen.
Wie schon bei einem Hosting System von Virtual Hosts, kann dadurch die Anzahl an öffentlichen IP-Adressen eingespart werden. Die Weiterleitung erfolgt dabei wiederum mithilfe des angefragten Domainnamens.

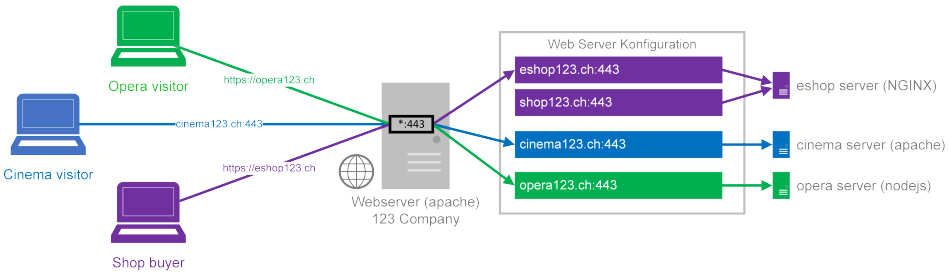


Abbildung 6.4: Vermittlung an Server hinter dem Reverse Proxy basierend der angefragten Webseiten

6.7 Reverse Lookup

Neben einem normalen DNS-Lookup, auch Forward DNS-Lookup genannt, bei welchem ein Hostname in eine IP-Adresse aufgelöst wird, kann mithilfe der IP-Adresse ein Reverse Lookup, oft auch Reverse DNS (rDNS) genannt, durchgeführt werden. Ein Reverse Lookup fragt in der DNS-Zone in-addr.arpa die gewünschte IP-Adresse ab und empfängt bestenfalls den Hostnamen eines Systems. Der Inhaber der IP-Adresse hinterlegt dafür den gewünschten Hostnamen, welcher bei einer Auflösung zurückgegeben werden sollte, in der DNS-Zone. Ein Hosting-Provider hinterlegt dafür meist den Namen des eigentlichen Hosting Systems.
Im Gegensatz zum normalen DNS-Eintrag wird der Reverse DNS-Eintrag weit weniger benötigt. Beispielsweise wird für die E-Mail-Kommunikation auf den Reverse Lookup als Spam Schutz zurückgegriffen, indem der E-Mail-Server die IP-Adresse des Senders von eingehenden E-Mails überprüfen. Stimmt der vorhandene Reverse Lookup nicht mit der Sender-Domain überein, wird die E-Mail als Spam klassifiziert. Der Reverse Lookup wird wie ein normaler DNS-Eintrag nicht vorgeschrieben, jedoch empfohlen [58].
Durch den Reverse Lookup ist es möglich den Hosting Provider einer Webseite zu ermitteln, da meist der Hostname des Hosting Systems oder Reverse Proxys zurückgegeben wird.

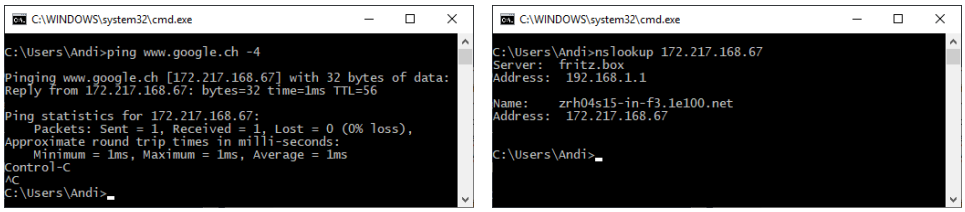


Abbildung 6.5: DNS Forward und Reverse Lookup für www.google.ch

6.8 TLS/SSL-Version

TLS kann für die Verschlüsselung der Datenübertragung von standardmässig unverschlüsselten Protokollen wie Hypertext Transfer Protocol (HTTP) oder SMTP verwendet werden. Als Nachfolger von SSLv3 mit der Versionsnummer 0x0300 besitzt TLS 1.0 die Versionsnummer 0x0301. Die aktuelle TLS-Version TLS 1.3 wird unter der Versionsnummer 0x0304 geführt [59].
SSLv3 gilt seit Oktober 2014 als nicht mehr sicher [60]. Aktuelle Browser und Systeme unterbinden darum seit Dezember 2014 standardmässig die Verbindung zu alten SSLv3-Seiten und setzen mindestens TLS 1.0 voraus [61]. Trotzdem können veraltete Systeme immer noch mit SSLv3 betrieben werden. So kann ein langjähriges Vernachlässigen von Softwareupdates, eine zu kostspielige Weiterentwicklung einer Software oder die teure Anschaffung von neuer Hardware zu solchen Problemen führen. Sofern die Geräte auch noch vernetzt sind, kann ein Angreifer um einiges einfacher Daten abgreifen, um diese danach für eigene Zwecke zu verwenden.

Die Aushandlung der genutzten TLS/SSL Informationen wird bei einer Erstverbindung mit einem Four-Way-Handshake ausgehandelt. [62]

- Client
  - **Client Hello**: Der Client startet die Kommunikation indem er dem Server seine unterstützten TLS- oder SSL-Versionen mitteilt. Zudem übermittelt der Client weitere für die Verschlüsselung relevante Informationen mit.
- Server
  - **Server Hello**: Der Server antwortet dem Client, nach einigen Überprüfungen auf seiner Seite, mit den für die zukünftige Kommunikation relevanten Informationen.
  - **Certificate**: Nach dem Server Hello sendet der Server dem Client zusätzlich noch sein signiertes Zertifikat und seinen öffentlichen Schlüssel zu.
  - **Server Key Exchange, Server Hello Done**: Der Server kann anhand des erhaltenen Client Hello nun den Key Exchange initiieren. Dabei sendet er die für den Client notwendigen Informationen zu. Im selben Paket sendet der Server noch das Server Hello Done, welches dem Client mitteilt, dass dieser nun weiterverfahren kann.
- Client
  - **Client Key Exchange**: Der Client generiert anhand der erhaltenen Server Informationen einen Schlüssel für die Verschlüsselung. Diesen sendet der Client dem Server, verschlüsselt mit dem öffentlichen Schlüssel des Servers, zu.
  - **Change Cipher Spec**: Der Client bestätigt danach, dass die weitere Kommunikation verschlüsselt stattfinden kann.
  - **Finished**: Der Client stellt dem Server das Finished zu. Diese Nachricht ist die erste verschlüsselte Nachricht auf dem Kanal.
- Server
  - **Change Cipher Spec**: Der Server bestätigt, nach dem Erhalt des Client Finished, dass er von nun auch verschlüsselt kommunizieren wird.
  - **Finished**: Zu guter Letzt bestätigt beendet der Server den Handshake mit einem Server Finished.

Die wichtigsten Informationen für die spätere verschlüsselte Verbindung, werden im ersten Client Hello und dem darauffolgenden Server Hello übermittelt. Nach dem Aufbau der verschlüsselten Verbindung, werden Informationen zur Verbindungsart nicht mehr nachgereicht.

6.9 Cipher Suite

Cipher Suites werden für die Verschlüsselung der Datenübertragung verwendet. Eine Cipher Suite setzt sich aus mehreren Informationen zusammen, wobei jeder Teil der Cipher Suite für einzelne Funktionen im Ablauf der Verschlüsselung zuständig ist. Die Cipher Suite legt somit das Vorgehen beim Verschlüsseln der Daten zwischen einem Client und einem Server fest. In der Abbildung 6.6 werden die darin enthaltenen Informationen aufgezeigt. [63]

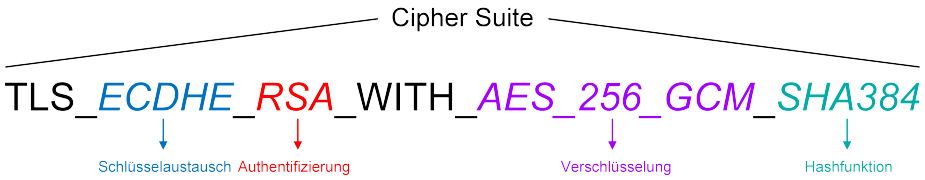


Abbildung 6.6: Einzelne Teile der TLS_ECDH_RSA_WITH_AES_256_GCM_SHA384 Cipher Suite

Die Aushandlung der Cipher Suite ist Bestandteil der Aushandlung einer TLS-Version. Der Client sendet dem Server in seinem Client Hello die von ihm unterstützten Cipher Suites zu. Der Server gleicht die empfangenen Suites mit den eigenen ab und sendet im Server Hello die von ihm als besten erachtete Cipher Suite, welche von beiden Seiten unterstützt wird, zurück. Anschliessend wird mit dieser Cipher Suite weiter kommuniziert. [59]
Cipher Suites können auch veraltet und dadurch nicht mehr empfohlen sein [30]. Sollte eine nicht mehr empfohlene Cipher Suite für die Kommunikation verwendet werden, könnte ein Angreifer die Kommunikation mit einer bekannten Schwachstelle der Cipher Suite entschlüsseln und in die übertragenen Informationen Einsicht erhalten. Die zu verwendenden Cipher Suites werden dabei auch von den jeweiligen Applikationen festgelegt. Veraltete Cipher Suites werden durch Software Updates oder Aktualisierungen der Server Konfigurationen verhindert, was bedeutet, dass Systeme immer auf dem aktuellsten Stand sein sollten [64] und auch die Konfiguration dieser Systeme kontrolliert werden soll [65].

6.10 Threats

Computer Threats verfolgen das Ziel in Systeme einzudringen. Es gibt hierzu verschiedene Ansätze. Threats können grundsätzliche andere Ziele verfolgen. Meist sind sie finanzieller Natur, wie die Geldschöpfung durch Erpressung oder den Verkauf von Informationen. Trotzdem gibt es immer noch einzelne Angriffe, welche explizit auf das unwiderrufliche Löschen von Daten aus sind.

6.10.1 Social Engineering

Social Engineering verfolgt den Ansatz, dass ein jeweiliger Endbenutzer eines Systems den Zugang auf ein System gewährt. Der Endbenutzer wird mithilfe von falschen Webseiten oder falschen E-Mails zu einer Handlung überredet, mit welcher er Zugriff auf seine Informationen ermöglicht. Unter diese Kategorie fallen beispielsweise Phishing E-Mails, welche auf eine falsche E-Banking- oder Firmen-Seiten verlinken. Auf diesen Seiten soll danach der jeweilige Benutzername mit dem dazugehörigen Passwort eingegeben werden. Die falsche Seite übermittelt dem Angreifer danach die eingegebenen Informationen, welche er für seine Zwecke weiterverwenden kann. [66]

6.10.2 Unerwünschte Software

Im Gegensatz zu Social Engineering, bei welchem der Endbenutzer seine Informationen freiwillig preisgibt, verfolgt der Angreifer mit unerwünschter Software ein anderes Ziel. Zur Kategorie von unerwünschter Software zählen lokal installierbare Malware oder Webseiten mit schädlichem JavaScript. Angreifer können dabei unterschiedliche Ziele verfolgen. Ransomware verschlüsselt beispielsweise die Speichermedien eines Systems mit einem, vom Angreifer zufällig ausgewählten, Schlüssel. Der Angreifer fordert anschliessend, ein Lösegeld, welches meist in einer CryptoCoin-Währung bezahlt werden soll. Dadurch soll die Anonymität des Angreifers gewahrt werden. Der Angreifer stellt in Aussicht, dass nach Begleichung der Forderung der Schlüssel zum Entschlüsseln der Daten ausgehändigt wird. Beispiele aus der nahen Vergangenheit zeigen jedoch, dass eine Bezahlung des Lösegelds keine erneute Entschlüsselung der Daten garantiert [67]. Melde- und Analysestelle Informationssicherung (MELANI) empfiehlt darum, kein Geld zu überweisen, da so das Netzwerk nur weiter gestärkt wird, da sich der Angreifer durch die erhaltenen CryptoCoins den benötigten Aufwand finanzieren kann. [68].

6.10.3 Entfernte Angriffe

Angreifer können sich aber auch von extern Zugang auf Systeme verschaffen. Ungeschützte oder nicht geupdatete Systeme sind dabei Hauptziele eines solchen Angreifers. Wenn der Zugang auf ein System möglich ist, wird versucht an erhöhte Rechte zu gelangen, mit welchen sich Programme ausführen lassen, welche zum Abhören von Daten oder zur Weiterverbreitung von Schadcode verwendet werden können.

7 Backend

Das Splunk Backend ist in einzelne Prozessschritte geteilt. Die Prozesse laufen dabei in vorgegebenen Intervallen oder zu speziellen Zeiten ab.

Name	**Prozess / Script**	**Eingabe**	**Ausgabe**	**Laufzeit**
Download	python3 traffic-analyzer.py *download*	CSV-Datei	Angepasste CSV-Datei	Intervall oder Cron-Job
Konvertierung	python3 traffic-analyzer.py *convert*	PCAP-Datei	Konvertierte CSV-Datei	Intervall oder Cron-Job
Enrichment	python3 traffic-analyzer.py *enrich*	Konvertierte CSV-Datei	Enrichte CSV-Datei	Nach Konvertierung
Monitor	splunkd	Enrichte CSV-Datei	Events in Splunk DB	Immer
API Endpoint	endpoint_handler.py	Key/Value	Angepasste .conf Dateien	Immer

Tabelle 7.1: Prozesse des Splunk Backends für die Splunk App Traffic-Analyzer

7.1 Download von Informationen

IANA und Institute of Electrical and Electronics Engineers (IEEE) stellen wichtige Datenquellen zur Verfügung. Es werden folgende Quellen für die Splunk App benötigt:

- MAC Hersteller: Da PCAPs keine Informationen zu MAC-Hersteller beinhalten, wird die offizielle Liste von IEEE verwendet [69]. Mit der frei verfügbaren Comma Separated Values (CSV)-Liste ist eine Zuordnung von MAC-Adresse auf MAC-Hersteller möglich.
- Cipher Suites: Cipher Suites werden in den Captures als Hexadezimal (HEX)-Werte abgelegt. Die HEX-Werte werden mit der Cipher Suites Liste von IANA

überprüft [30]. Die Liste beinhaltet auch das Attribut, ob die jeweilige Cipher Suite als noch genügend sicher klassifiziert wird.

Diese Quellen werden, wenn die Default-Werte vom Benutzer nicht überschrieben werden, stündlich heruntergeladen, angepasst und in einem vordefinierten Ordner abgelegt. Dadurch, dass sich die Listen über die Zeit weniger ändern, wäre ein grösseres Updateintervall möglich. Ein kleineres Update-Intervall würde einen weit grösseren Overhead, mit schon vorhandenen Informationen, mit sich bringen.

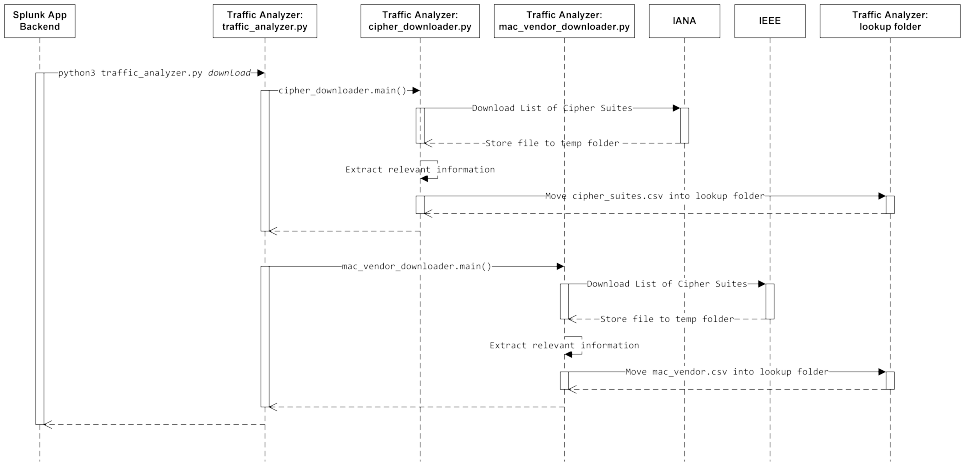


Abbildung 7.1: Sequenzdiagramm des Download Prozesses

7.2 Konvertierung

In der Konvertierungsphase werden die PCAP-Dumps mithilfe von Tshark konvertiert. Hierzu werden die zu benötigten PCAP-Dateien mithilfe von Tshark eingelesen und die Ausgabe auf der Konsole in eine Datei umgeleitet. Durch die kommaseparierte Trennung der Felder kann somit eine CSV-Datei erstellt werden.
Vor der eigentlichen Konvertierung mit Tshark wird von den jeweiligen PCAPs der Hashwert errechnet. Der errechnete Hashwert wird anschliessend in einer Liste mit Hashwerten von schon konvertierten PCAP-Dateien gesucht. Sofern der Hashwert nicht in der Liste vorkommt, fährt man mit dem Konvertierungsprozess fort. Anderenfalls wird die PCAP-Datei übersprungen.
Dabei werden nur vereinzelte Attribute eines Packets exportiert. Tshark bietet auch die Möglichkeit einzelne Attribute zu errechnen. Solche Attribute gehören den Display Filter Referenzen _ws. Attributklasse an.
Für das Dashboard Protocols wird dabei das Attribut _ws_col_Protocol genutzt. Dieses zeigt dabei das am höchsten, innerhalb des Open Systems Interconnection (OSI)-Layer, detektierte Protokoll eines Packets an. Nach der Konvertierung einer PCAP-Datei wird der errechnete File Hash abgespeichert. Dadurch werden nur Leserechte auf dem PCAP-Ordner benötigt. Der PCAP-Ordner wird über das Konfigurationsfile der App defi-

niert und im Backend eingelesen.

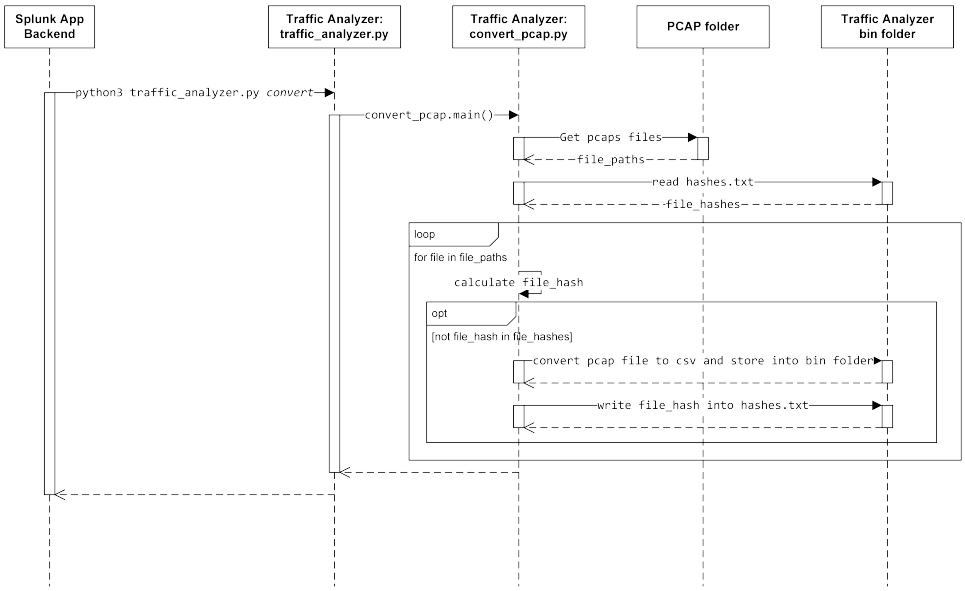


Abbildung 7.2: Sequenzdiagramm der Konvertierung von PCAPs

Nach der Konvertierung wird die neue CSV-Datei dem Enrichment übergeben.

7.3 Enrichment

Das Enrichment reichern die neu erhaltenen CSV-Dateien mit weiteren Informationen an. Dabei läuft das Enrichment immer nach dem Konvertierungsprozess an. Der Prozess des Enrichments analysiert dabei Zeile um Zeile der CSV-Dateien. Jede Zeile steht dafür für ein einzelnes Packet eines Captures. Einige Informationen zu einer Verbindung tauchen dabei nur einmalig auf, sind aber für den weiteren Verlauf von Relevanz.

7.3.1 Server Funktionen

Für das Dashboard Server Types 5.5.2 werden bestimmte Flags pro Server Typ im Enrichment Prozess gesetzt. Dabei werden die untenstehenden Server Funktionen unterschieden.

DHCP Server

Anhand von DHCP Offers wird erkannt, dass die Source IP ein DHCP-Server ist. Sofern ein DHCP-Server erkannt wird, wird das Flag *is_dhcp_server* gesetzt.

DNS Server

Anhand der DNS standard query response wird ein DNS-Server erkannt. Das Packet wird mit dem gesetzten Flag *is_dns_server* gekennzeichnet.

7.3.2 Stream ID

Obwohl Netzwerksniffer wie Wireshark in ihren Captures eigene Stream IDs definieren, wird eine eigene Stream ID pro Verbindung errechnet. Netzwerksniffer wie Wireshark errechnen die Stream ID pro Capture. Dies hat zur Folge, dass bei mehreren, aufeinanderfolgend durchgeführten Captures, die gleiche Stream ID für verschiedene Verbindungen definiert werden würde.
Die eigens errechnete Stream ID wird dabei anhand der Source und Destination Sockets, sowie dem Protokolltyp der jeweiligen Verbindung definiert. Dabei werden die Informationen so kombiniert, dass die gleiche Stream ID bei ausgehendem, sowie eingehendem Traffic ermittelt wird. Die neue Stream ID ist dabei ein numerischer Wert, welcher durch das Hashing und anschliessende Modulieren bestimmt wird.
Die neue Stream ID kann somit über mehrere Captures verwendet werden.

7.3.3 Geo Standort

API

Für die Ermittlung eines IP-Standortes wird dabei auf den Online-Dienst von KeyCDN zurückgegriffen [42]. Durch die vorhandene API und die vergleichsweise hohe Anzahl an Aufrufen pro Sekunde [1] [2] [3] ist KeyCDN der optimalste Anbieter. Lizenztechnisch wird eine Verlinkung im Repository vorgeschrieben, welche mithilfe eines Badges vorgenommen wird.

[1] KeyCDN: 3 Anfragen pro Sekunde [42]
[2] ipgeolocation: 0.011 pro Sekunde [44]
[3] ipstack: 0.003 Anfragen pro Sekunde [43]

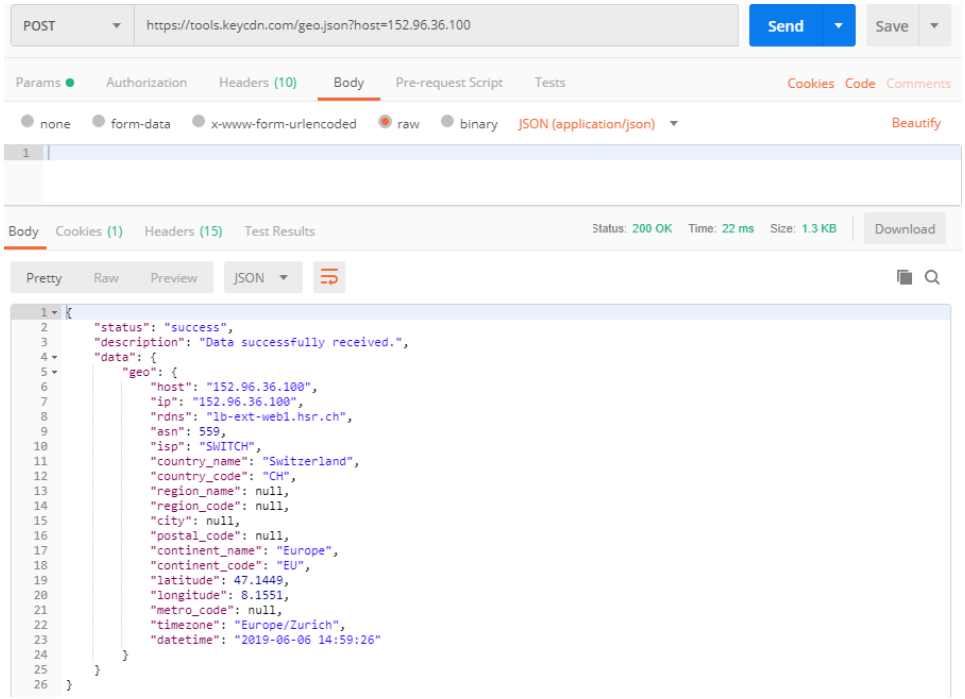


Abbildung 7.3: Antwort der API von KeyCDN für die IP-Adresse 152.96.36.100

KeyCDN sendet bei der Anfrage an die JavaScript Object Notation (JSON)-API ein Objekt mit den benötigten Informationen zurück. Für die Bestimmung des Geo-Standortes sind dabei die Attribute data.geo.latitude und data.geo.longitude von Relevanz. Weitere Informationen, welche den Standort definieren, werden für die Lokalisierung nicht benötigt, da diese anhand der Längen- und Breitengrade bestimmt werden können.

Clientseitige Ortungslösung

Für die Arbeit wurden clientseitige Ortungslösungen nicht weiterverfolgt, da die Analyse ohne Unterstützung der Clients stattfinden sollte. Bei einer späteren Lokalisierung eines Clients kann auch nicht davon ausgegangen werden, dass sich der Client noch am selben Ort wie zum Messzeitpunkt befindet.

Standard Standort

Bei der Lokalisierung einzelner Systeme greift KeyCDN auf Standard Standorte zurück. Obwohl der Hostname *zrh04s06-in-f131.1e100.net* suggeriert, dass sich der Server in der Region Zürich befindet, returniert KeyCDN, dass der Server in den USA stationiert wäre.

Abbildung 7.4: Falsche Standortauflösung des Servers *zrh04s06-in-f131.1e100.net* mithilfe der IP Location Finder Webseite von KeyCDN [42]

Solche False Positives existieren dabei nicht nur bei KeyCDN, auch ipstack liefert solche Werte über ihre API zurück [70]. Andere Anbieter wie ipgeolocation [71] returnieren dabei einen Standort, welcher noch am ehesten an den effektiven Standort des Servers herankommt.

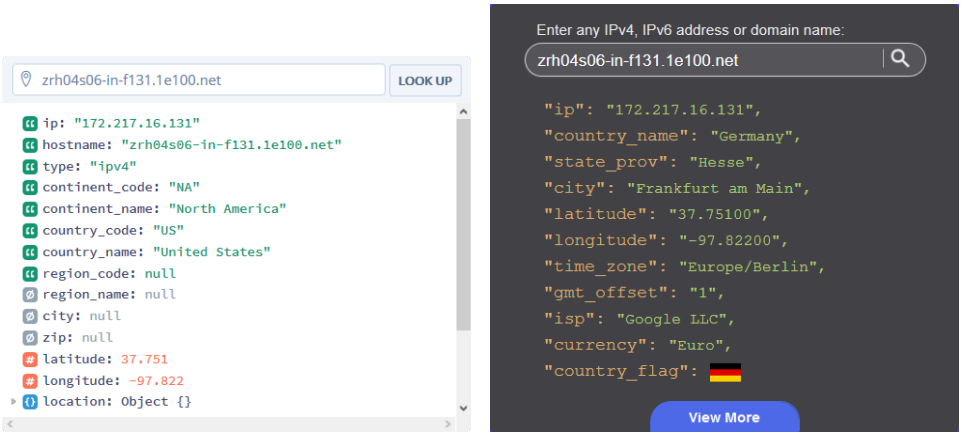


Abbildung 7.5: Falsche (links, ipstack [70]) und richtige (rechts, ipgeolocation [71]) Standortauflösung des Servers *zrh04s06-in-f131.1e100.net* mithilfe von IP Geolocation Diensten

Die Nutzung solcher falschen Informationen könnte mithilfe von zuvor im Kapitel 6.4.2 erwähnten Ping Tests Tools unterbunden werden. Aber auch durch die Nutzung von Ping Tests, könnte man nicht bestimmen, ob das Gerät nun wirklich am richtigen Standort lokalisiert wurde. Auch wenn die Lokalisierung genau genug wäre, würde die Überprüfung für jede einzelne öffentliche IP innerhalb eines Captures einen unrentablen Overhead produzieren.
Für die Entwicklung wird dadurch auf die Überprüfung der erhaltenen Information verzichtet.

7.3.4 Reverse Lookup

Zur Erkennung des Hosting Providers wird der automatisch mitgelieferte Reverse Lookup von KeyCDN genutzt. Das Attribut data.geo.rdns returniert dabei die Domain. Durch die Nutzung der API, anstelle der clientseitigen Auflösung, kann pro öffentliche IP-Adresse ein zusätzlicher DNS Reverse Lookup Request eingespart werden. Einzelne öffentliche Adressen landen zudem in einem DNS request timeout, welcher anderenfalls bei einer clientseitigen Auflösung zwei Mal auftritt. Einmal bei der KeyCDN-Abfrage und ein weiteres Mal bei der lokalen Auflösung.

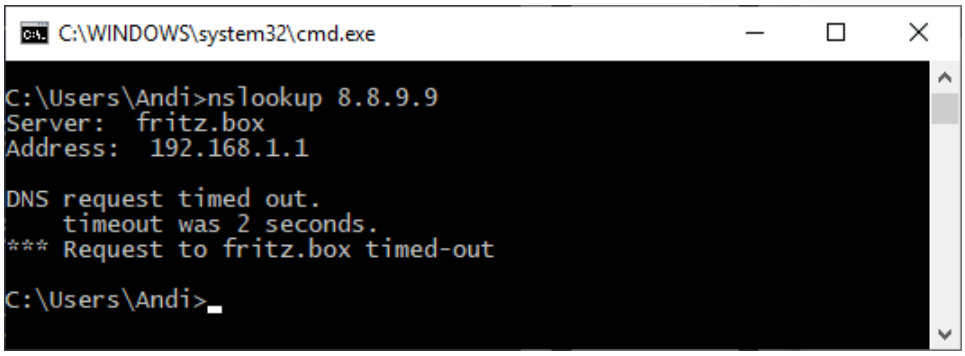


Abbildung 7.6: DNS-Timeout bei einer Reverse Lookup Anfrage für die IP-Adresse 8.8.9.9

Für interne Adressen wird der Reverse Lookup über den DNS-Server des Splunk Systems vorgenommen. Zusätzlich werden, sofern vom Benutzer so konfiguriert, zusätzliche DNS-Server über die Konfigurationsdatei verwendet, sofern diese erreichbar sind.

7.3.5 DNS Erkennung

Da der Reverse Lookup meist nicht den eigentlich angefragten Hostnamen zurückliefert, werden die DNS-Response Pakete abgefangen. Anhand des jeweiligen DNS-Response wird die zurückgegebene IP-Adresse mit dem anschliessend aufgebauten Stream gemapped. Somit kann dem Stream nicht nur der Reverse Lookup sondern auch der eigentliche Forward Lookup zugeordnet werden.

7.3.6 TLS/SSL-Version

Die ausgehandelte TLS/SSL-Version wird bei jedem Server Hello aufgezeichnet und der jeweiligen Stream ID zugeordnet. Für spätere Zwecke wird auch in jedes zukünftige Packet eines Streams die TLS/SSL-Version geschrieben werden.

7.3.7 TLS verschlüsselter Traffic

Durch das Attribut _ws_col_Protocol können Packets zwar als TCP-Packets detektiert werden, auch wenn sie einer TLS-Verbindung angehören. Dies kommt daher, dass die Packets segmentiert werden.

7.3.8 Cipher Suite

Wie schon die TLS/SSL-Version wir die Cipher Suite nur beim Aufbau der verschlüsselten Verbindung übermittelt. Gleich wie die TLS/SSL-Version wird diese auf ein Stream zugewiesen und für die Referenz für die weitere Packets der Verbindung genutzt.
Die Ermittlung, ob die Cipher Suites noch empfohlen wird, wird im Frontend abgehandelt.

7.3.9 Webserver Typen

CDN Server

WebPageTest stellt für die Erkennung von CDN-Servern eine Liste in ihrem GitHub Repository zur Verfügung [72]. Das vorhandene Dictionary *OptimizationChecks.cdn_cnames* wird dazu identisch übernommen.
Sofern der DNS Name der Source Adresse in dem Dictionary gefunden wird, wird das Flag *is_cdn_server* gesetzt.

Social Network Server

Für die Erkennung von Social Network Servern wird die zusammengesetzte Liste von StevenBlack genutzt [73]. Aus der Liste werden die Domains so aggregiert, dass ein Testen auf Sub- und Hauptdomain durchgeführt werden kann.
Wie schon bei der Erkennung eines CDN-Servers wird ein Social Netzwork Server anhand der Source Domainnamens erkannt. Dabei wird gegen die neu erstellte Social Network Liste getestet. Sofern der Domainname in der Liste vorkommt, wird das Flag *is_social_network_server* gesetzt.

7.3.10 Threat Erkennung

Threats werden mithilfe der Google Safe Browsing API detektiert. Die Google Safe Browsing API bietet dabei eine kostenlose Threat Detection API an. Für die API wird einzig ein API-Key benötigt, welcher über die Google Developer Console erstellt werden kann. Die Safe Browsing API unterscheidet dabei zwischen fünf Threat Types. [74]

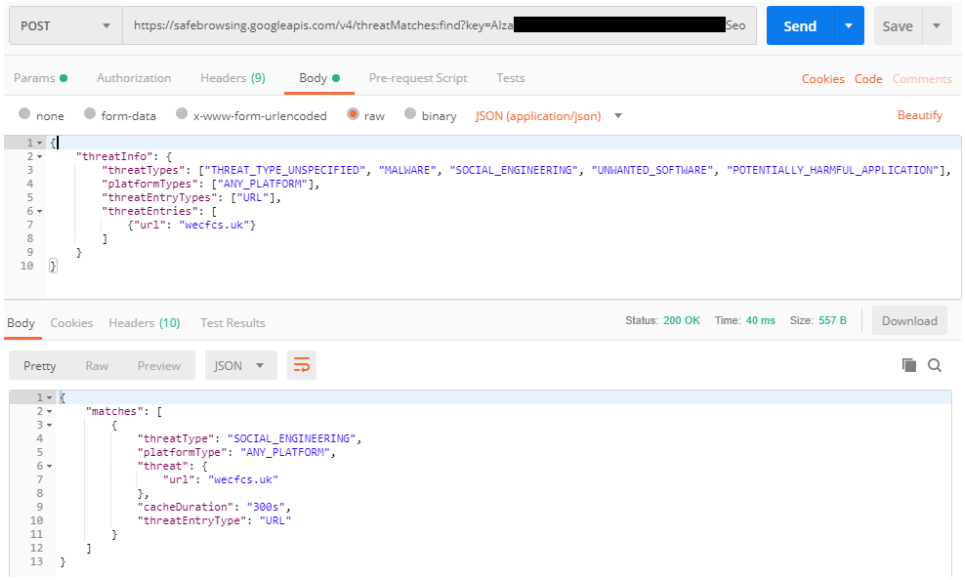


Abbildung 7.7: Antwort der Google Safe Browsing API für den Domänennamen wecfcs.uk

Der Request Body wird für die API-Abfrage mit den Hostnamen einer DNS Response befüllt. Dadurch wird überprüft, ob der später angefragte Server ein Threat ist. Sofern die API einen Threat Typ zurücksendet, werden diese numerisch, als zusammengesetzte, kommaseparierte Werte in das neue CSV geschrieben.
Es können dadurch nur Threats basierend auf Domainnamen erkannt werden. Threats, wie beispielsweise Distributed Denial of Service (DDOS) werden dadurch aber nicht erkannt.

7.3.11 Ad/Tracking Detektierung

Durch die vorhandenen DNS-Informationen können die angefragten Domainnamen gegen eine Liste von Ad und Tracking Server getestet werden. Die herzu genutzte Liste wird von Yoyo.org heruntergeladen und entsprechend geparsed [75]. Sofern die angefragte Domain oder Subdomain in der Liste auftaucht, wird der aktuelle Stream als Ad oder Tracking Traffic mithilfe des Flags is_ad_or_tracking deklariert. Da die von Tracking Anbietern gesetzten Cookies oder übermittelten Headers per HTTPS übermittelt werden, ist eine Unterscheidung zwischen Tracking und Ad, ohne entschlüsselten Netzwerkverkehr oder weitere Informationen des Browsers, nicht möglich.

7.4 Monitoring

Das Splunk Backend überwacht einzelne Order auf veränderte Dateien. Sofern sich eine Datei ändert oder eine neue Datei im Ordner landet, wird diese Datei automatisch in die Splunk Datenbank eingelesen. Die Monitor Jobs werden dabei in der Datei *inputs.conf* konfiguriert. Die Konfiguration definiert pro Monitor einen Source Type. Mit dem key/va-

lue pair «crcSalt = <SOURCE>» wird festgelegt, dass die Datei nur dann neu eingelesen wird, wenn sich der Hash der Datei ändert. [76]

```
[monitor://$SPLUNK_HOME/etc/apps/traffic-analyzer/lookups/captures/]
disabled = 0
sourcetype = capture
crcSalt = <SOURCE>

[monitor://$SPLUNK_HOME/etc/apps/traffic-analyzer/lookups/lists/]
disabled = 0
sourcetype = list
crcSalt = <SOURCE>
```

Abbildung 7.8: Definition der Monitors in der Datei inputs.conf

7.4.1 Source Type

Source Types werden in Splunk genutzt, um Events zu kategorisieren. Source Types können hierzu selbst erstellt werden. Beim Einlesen von neuen Events wird der Source Type definiert. Source Types werden dabei in der Datei *props.conf* definiert. [77]

```
[capture]
disabled = false
description = Props for traffic-analyzer enriched csv files
SHOULD_LINEMERGE = false
FIELD_DELIMITER = ,
HEADER_FIELD_DELIMITER = ,
FIELD_QUOTE = "
```

Abbildung 7.9: Source Type Definition von einzulesenden Captures in der Datei props.conf

7.5 API Endpoints

Für die Einstellungen der Splunk App wird ein eigener API Endpoint erstellt. Das Dashboard Einstellungen übermittelt die eingetragenen Daten dem Endpoint, welcher die empfangenen Daten in die Konfigurationsdateien schreibt. Dafür wird ein Python Skript erstellt, welches die eingehenden Anfragen bearbeitet. Das Skript wird auf der Vorlage von Splunk entwickelt [78].

8 Frontend

8.1 Dashboards

Splunk Dashboards werden mit der Simple XML Notation von Splunk erstellt [79].

8.1.1 Queries

Zugriffe auf Daten in Splunk werden über die SPL umgesetzt. Mit wachsenden Datenmengen und volleren Dashboards wurden die Abfragen jedoch zunehmend langsamer. Um die in Tabelle 3.1 gelisteten NFRs betreffend der Performance (# 6 und # 7) dennoch einhalten zu können, mussten die Abfragen mithilfe der Post Process Search optimiert werden.

Post Process Search

Durch die Nutzung von Post Process Searches werden Base Seaches definiert, welche die Daten an weitere Queries weiterreichen. Dadurch ist das Auslagern von Abfrageschritten möglich, was einen positiven Einfluss auf die Performance hat. [80]

8.1.2 Drilldown

Auf zwei Dashboards werden spezielles Drilldown-Verhalten festgelegt, um weiterführende Informationen anzeigen zu können. So wird mit einem Klick auf ein Kuchendiagram auf den Dashboards Overview (Kapitel 5.5.1) und Protocols (Kapitel 5.5.5) der Wikipedia-Artikel für das angeklickte Protokoll geöffnet. Durch Beschränkungen in den Möglichkeiten beim definieren solcher Drilldowns funktioniert ein solcher Aufruf nur, wenn der Name des Artikels auf Wikipedia effektiv dem im Dashboard angezeigten Namen des Protokolls entspricht.

8.1.3 Datenschutzbestimmungen einzelner Länder

Als Liste der Datenschutzbestimmungen einzelner Länder werden die vom EDÖB verfügbaren Länderinformationen manuell in ein CSV geschrieben und in Splunk einmalig eingelesen [31].

8.2 Cascading Style Sheets (CSS)

Für die bessere Darstellung von einzelnen Eingabefelder der Filter wird ein eigenes Cascading Style Sheets (CSS) geschrieben. Das CSS wird dabei mithilfe von Sass in der Sassy CSS (SCSS)-Syntax definiert. Die erstellte SCSS-Datei wird dabei von Sass in eine CSS-Datei kompiliert. Sass bietet dabei den grossen Vorteil Definitionen Auszulagern und wiederzuverwenden. Dadurch ist ein einfacheres Erstellen von CSS-Dateien möglich. [81]

8.3 Tokens

Splunk nutzt Tokens für die Übertragung von Informationen über mehrere Suchen [82]. Tokens werden dabei standardmässig nicht über mehrere Dashboards hinweg genutzt. Tokens können standardmässig nicht über die Dashboards einer App hinweg genutzt werden.
Für die genutzten Filter der Splunk App wird jedoch ein JavaScript geschrieben, welches die einzelnen Tokens, sobald sie gesetzt werden, in den SessionStorage gespeichert [83]. Bei jedem neuen Seitenaufruf einer aktuellen Session, werden die Tokens aus dem SessionStorage genommen und wieder als Tokens gespeichert, welche das jeweilige Dashboard anschliessend nutzen kann. So sind gleiche Filter über mehrere Dashboards anwendbar, ohne dass der Nutzer diese immer wieder setzen muss.

8.4 Usability Test

Um sicherzustellen, dass die Applikation nicht an den Bedürfnissen von möglichen Nutzern vorbei entwickelt wird, werden Usability Tests durchgeführt. Die Resultate dieser Tests sind im Anhang zu finden.
Es zeigt sich, dass die Applikation bei den Nutzern in den Tests weitgehend gut ankommt. Es wird jedoch gewünscht, dass durch die Applikation an Orten, wie beispielsweise dem Dashboard Protocols, Erklärungen zu den dargestellten Informationen anbietet. Ein Lösungsansatz dazu wird in Kapitel 8.1.2 aufgezeigt.

9 Entwicklung

9.1 Eingesetzte Entwicklungs-Tools

Für die Entwicklung eines Splunk Apps werden folgende Tools verwendet.

9.1.1 Python

Zur Entwicklung der App wird Python 3.7.3 verwendet. Da Splunk selbst noch auf Python 2.7.x basiert, muss Python zusätzlich installiert sein, da geschriebener Python Code in einer Version nicht mit der anderen Version kompatibel sind. Python 2.7 dabei die letzte Version unter der Hauptversion 2.x. Python 2.7 geht am 1. Januar 2020 End-Of-Life. Danach wird der kostenlose Support eingestellt. Entwicklern wird darum schon länger empfohlen nach Python 3.x zu migrieren. [84]

9.1.2 Integrated Development Environment (IDE)

Als Integrated Development Environment (IDE) wird PyCharm verwendet. PyCharm unterstützt neben Python noch weitere Programmiersprachen. PyCharm liefert auch weitere Features wie beispielsweise einen Profiler oder eine integrierte Code Analyse mit. Der Profiler ermöglicht es Funktionen auf ihre Laufzeit zu analysieren und somit gezielte Verbesserungen der Performance zu ermöglichen. Mit der integrierten Code Analyse sind Verletzungen des Python Enhancement Proposal (PEP) 8 schnell lokal einsehbar.

9.1.3 Source Control Management (SCM)

Für das Source Code Management (SCM) wird GitHub verwendet. Das Repository ist öffentlich verfügbar und gehört der für die Bachelorarbeit erstellten GitHub Organisation anjo-hsr an. Das Repository ist unter dem Link `https://github.com/anjo-hsr/Traffic-Analyzer` zu finden. Das GitHub Repository ist dabei mit einzelnen Services verbunden.

9.1.4 Continous Integration und Continous Deployment (CI/CD)

Für Continuous Integration (CI) wird Travis CI eingesetzt. Travis CI ist für Open Source Projekte gratis. Bei Travis CI können mehrere Build Steps hinterlegt werden, welche verschiedene Tasks, abhängig voneinander durchführen. Travis CI läuft mit jedem Push

auf einen Branch an. Sofern irgendein Task von Travis CI fehlschlägt, gilt der Build als fehlgeschlagen. Ein fehlgeschlagener Build kann dabei nicht in den master Branch gemerged werden.
Es werden zwei unterschiedliche Steps zum Continuous Deployment (CD) ausgeführt. Beide Schritte werden nur mit einem Push auf den Master Branch ausgeführt. Jenkins führt auf einem Rechner der HSR einen Job aus, welcher den aktuellen Master Branch als Docker Container auf dem System ausführt. Mithilfe von Travis CI wird jeweils ein Docker Image gebuildet, welches an Docker Hub übermittelt wird. So können andere potenzielle Nutzer der App, diese einfacher mit Docker deployen.

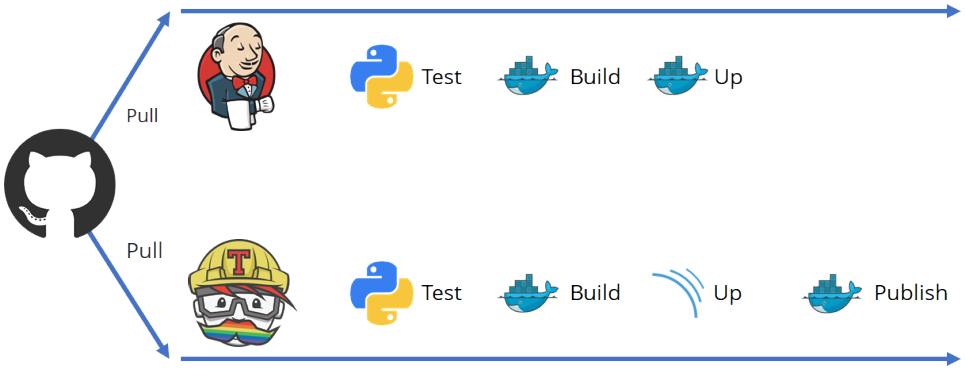


Abbildung 9.1: Continous Deployment von Traffic-Analyzer

9.1.5 Code Quality

Die Code Quality wird mit SonarCloud gewährleistet. SonarCloud bietet eine statische Codeanalyse an, welche die gängigsten Code Smells, Vulnerabilities und Bugs identifizieren kann. SonarCloud zeigt auch die aktuelle Code Coverage der Tests für einen Branch an.
SonarCloud definiert dabei einzelne Schwellwerte, welche aufzeigen, wie wartbar oder sicher der analysierte Code ist. Sofern diese überschritten werden, ist die Code Quality nicht mehr gegeben und das mergen auf den master Branch wird unterbunden.
Travis CI startet beim zweitletzten Build Step einen Scan für die Code Quality und übermittelt die resultierenden Daten an SonarCloud.

9.1.6 Testing

Das Backend der Splunk App wird mithilfe von Coverage.py getestet [85]. Coverage.py liefert dabei die Ergebnisse der Unit Tests und die Code Coverage der einzelnen Python Files. Travis CI nutzt Coverage im ersten und zweitletzten Build Step. Im ersten Build Step wird darauf geachtet, ob die Unit Tests nach einem Commit immer noch erfolgreich sind. Sofern ein Test fehlschlägt, wird der Build des Commits abgebrochen und der Build schlägt fehl.
Sofern die Unit Tests erfolgreich waren, werden die Statistiken der Code Coverage im letzten Build Step an SonarCloud weitergereicht.

9.1.7 Linter

Für die Python Programmiersprache wird zusätzlich zu den SonarCloud gegebenen Mitteln der Linter PyLint bei jedem Build auf Travis CI gestartet. PyLint liefert SonarCloud zusätzliche Informationen zum PEP 8 Standard [86].

9.1.8 Containerization

Containerization ist ein alternativer Virtualisierungsansatz gegenüber der Hardware-Virtualisierung. Im Gegensatz zur Hardware-Virtualisierung nutzt die Container-Virtualisierung den vom Host-System verwendeten Kernel. Zugriffe auf die Hardware finden direkt statt. Durch die Abstraktion auf der Anwendungsschicht ist der Overhead für die Nutzung einer Applikation weit kleiner, da die Installation von einem virtuellen Betriebssystem mit allen Treibern und Systemdateien entfällt. Für die Entwicklung wird die Containerization-Lösung Docker verwendet.

Container

Container sind virtuelle Hosts einer Containerization-Lösung. Container laufen dabei in einem jeweils getrennten Prozess innerhalb des User Spaces. Der User Space ist der Teil des Memories, welcher keine Kernel-Prozesse ausführt. Die Trennung der Prozesse ermöglicht eine Abgrenzung der einzelnen Hosts. Container laufen in einer Art Sandbox, welche unerwünschte Zugriffe zwischen Containern und dem Hostsystem unterbindet. Für die Splunk-App Entwicklung wird ein eigener Container verwendet. Innerhalb des Containers läuft eine Splunk-Instanz.

Images

Images sind Templates zur Erstellung von Containern. Sie werden mithilfe von bestehenden, konfigurierten Containern erstellt und können dadurch selbst erstellt oder von einem öffentlichen Image Repositories, wie beispielsweise Dockerhub, heruntergeladen werden. Images ermöglichen einen vereinfachten Umgang mit Containern, da viele, unterschiedliche Services über Image Repositories bereits zum Download zur Verfügung stehen. Mithilfe von Environmentvariabeln oder Konfigurationsdateien können Images weitgehend fertig konfiguriert oder angepasst werden. Dadurch ist eine weitere Anreicherung eines heruntergeladenen Images möglich. Um ein schnelleres Deployment zu ermöglichen wird vom Container der Splunk-App-Entwicklung ein Image erstellt.

Docker

Docker ist eine Open Source Containerization-Lösung, welche im Jahr 2013 veröffentlicht wurde. Docker bietet eine einfache Handhabung von Containern mithilfe von sogenannten Dockerfiles an. Dockerfiles beinhalten die Konfigurationsschritte, um ein Image zu erstellen. Dabei können Ordner kopiert, Software nachinstalliert oder weitere Services konfiguriert werden.

Docker bietet mit Docker-Compose zusätzlich die Möglichkeit das Erstellen und Starten von mehreren Images innerhalb einer Datei zu definieren. Dabei ist es möglich Abhängigkeiten einzelner Container festzulegen, virtuelle Netzwerke zu erstellen oder Ports an den Container weiterzuleiten. Mit Docker wird die Splunk App Entwicklung auf allen Entwicklungsgeräten identisch gehalten. Dank Docker ist eine individuelle, aber trotzdem einfach zu bedienende Testumgebung möglich. Durch das einfache Löschen und Erstellen eines Containers, hält sich der administrative Aufwand klein.
Docker wird neben den Entwicklungsgeräten auch bei Travis CI eingesetzt. Zudem kann das produktive Splunk-System auch als Docker Container genutzt werden.

9.1.9 Tshark

Für die Konvertierung von PCAP-Captures in CSV-Dateien, wird auf den Hauptsystemen Tshark in der Version 3.0.2 verwendet. Der Docker Container von Splunk basiert auf einem Debian, welches, bis zum Stand der Abgabe, über den Package Manager apt nur Tshark in der Version 2.6.7 zur Verfügung hatte [87].
Wireshark nutzt seit der Version 3.0.0 jedoch andere Display Filter [1] [2], auch dissectors genannt [88]. Zwar werden die alten Filternamen noch unterstützt, trotzdem werden in den Skripts standardmässig die neuen Namen angegeben. Um eine Rückwärtskompatibilität zu gewährleisten, wird mithilfe von stream editor (sed) ein Regex ausgeführt, welcher in die den Skripts definierten Filternamen mit den alten Bezeichnungen ersetzt [89].
Für die spätere Nutzung in den Frontend Queries wird beim Enrichment der Header der Dateien wieder zu den neuen Bezeichnungen umbenannt, sodass die Queries mit den neuen Namen aufgebaut werden können [90].

9.2 Deployment Diagramm

Für die Entwicklung der Splunk App Traffic-Analyzer wird das folgende Deployment Diagramm umgesetzt.

[1] SSL neu TLS
[2] BOOTP neu DHCP

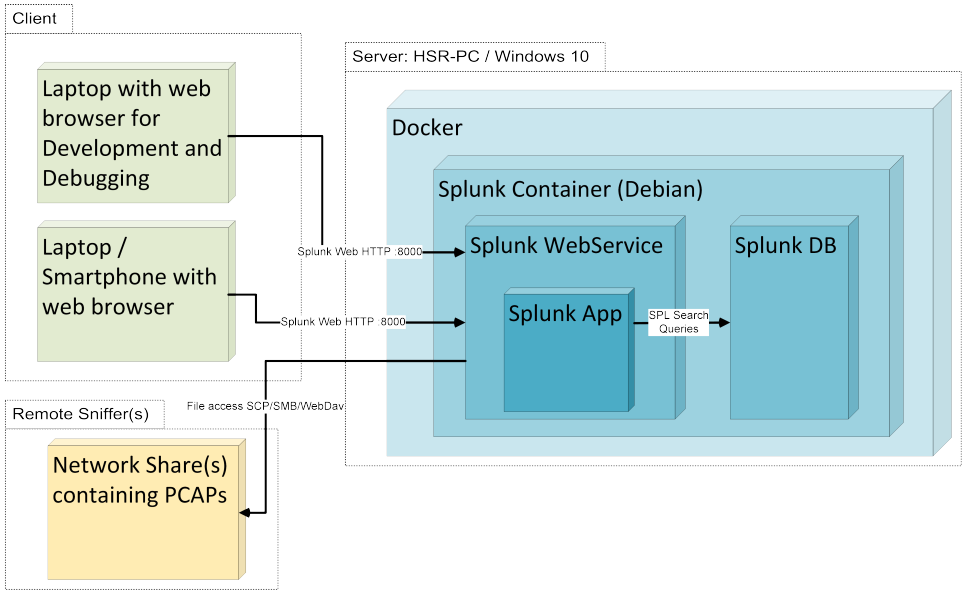


Abbildung 9.2: Deployment Diagramm von Traffic-Analyzer

10 Nutzung

10.1 Betriebssystem

Neben der Nutzung mithilfe eines Docker Containers ist die Splunk App nur mit einer Splunk Instanz auf einem Linux Systemen nutzbar. Letzte Anpassungen für eine Splunk Installation auf einem Windows Rechner, konnten nicht mehr umgesetzt werden.

10.2 Dependencies

Für die Entwicklung der Splunk App werden einzelne Python Dependencies benötigt. Diese sind unter der Datei requirements.txt zu finden. Die jeweilige Dependency muss auf dem System, welches Splunk ausführt, installiert werden.
Zusätzlich zu diesen Dependencies wird Tshark und eine Python 3.x Umgebung benötigt.

10.3 Lizenz

Die entwickelte Splunk App steht unter der MIT-Lizenz [91]. Dabei sind die genutzten Dependencies zur Lizenz kompatibel.

10.4 Docker Hub Image

Traffic-Analyzer lässt sich einfach mithilfe vom veröffentlichten Image auf Docker Hub `https://hub.docker.com/r/anjohsr/traffic-analyzer` deployen. Hierzu wird lediglich eine Docker Installation benötigt, mit welcher sich dann das Image ausführen lässt.

11 Messaufbau

Für die Analyse von einzelnen Webseiten oder Smart Home Devices werden von folgenden Messaufbauten genutzt. Für die Messungen, welche später für die Analyse ausgewertet werden, wird ein Allegro Network Multimeter der Firma Allegro Packets verwendet. Die darauf installierte Firmware ist die Version 2.3.0.
Bei jeder Messung wird sichergestellt, dass zeitgleich nur ein Device mit dem Router verbunden ist. Die anderen Devices sind zum Zeitpunkt der Messung entweder vom Strom abgehängt oder ausgeschaltet.
Die PCAP-Dateien werden auf einer externen Festplatte gespeichert. Die Daten werden nach jedem erfolgten Test auf das System mit einer laufenden Splunk Instanz und installierter Splunk App kopiert.

11.1 Messaufbau für kabellose Verbindungen

Die Messung von kabellosen Verbindungen wird zwischen dem Router und der Firewall vorgenommen, da eine Spiegelung des WLAN-Verkehrs, auf einen anderen Routing Port, nicht möglich ist.

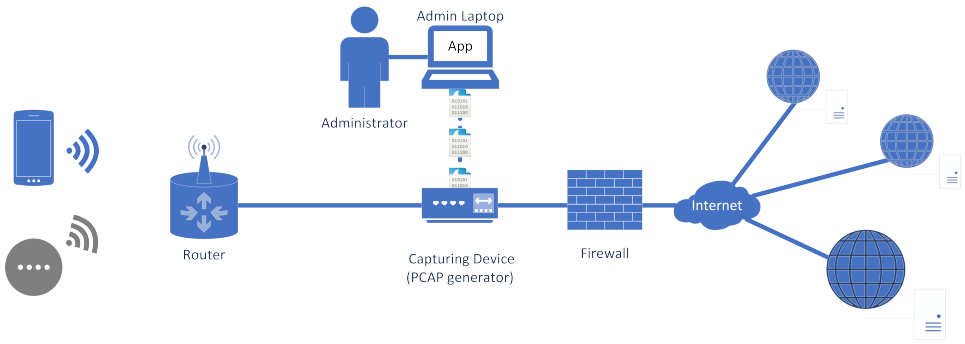


Abbildung 11.1: Messaufbau für das Analysieren von kabellosem Verkehr

11.2 Messaufbau für kabelgebundenen Verbindungen

Für die Messung von kabelgebundenen Verbindungen wird das Aufnahmegerät zwischen dem effektiven Gerät und dem Router zwischengeschaltet.

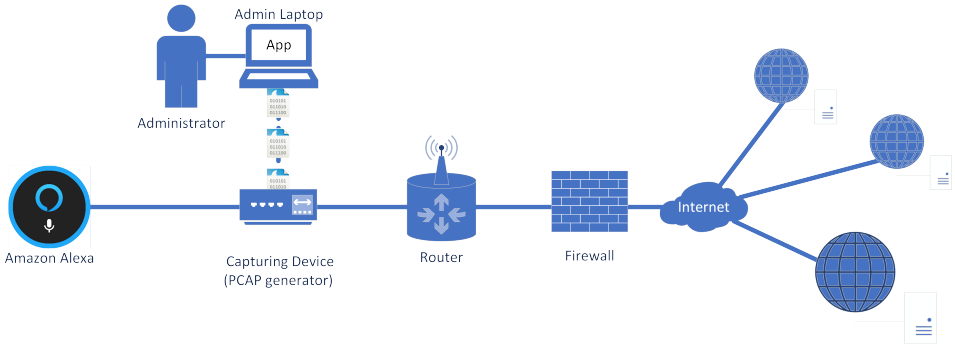


Abbildung 11.2: Messaufbau für das Analysieren von kabelgebundenem Verkehr

12 Analyse

Um die entwickelte Splunk App zu testen und um aufzuzeigen, was mit der App gemacht werden kann, werden im folgenden Abschnitt einige bekannte Mobile Apps, Webseiten und Speedtest analysiert. Das Setup der erstellten Messungen entspricht dem Setup aus Abbildung 11.1, wobei das für die Messung angeschlossene Endgerät jeweils ein anderes sein kann.
Alle Tests mit einem Smartphone wurden auf einem Sony Xperia Z5 Compact mit Android 7.1.1 durchgeführt.

12.1 SBB Mobile App

Kennzahl	**Wert (% von Total)**
Total number of streams	50
Number of inadvisable cipher suite sused	0 (0%)
Amount of unsecured traffic (by # of packets)	314 (34.13%)
Traffic incoming (Bytes)	153.91 KB
Traffic incoming (Packets)	395
Traffic outgoing (Bytes)	219.35 KB
Traffic outgoing (Packets)	481

Tabelle 12.1: Kennzahlen der Auswertung für die SBB Preview Mobile App

Für die Tests mit dem SBB Mobile App wurde die SBB Mobile Preview App [92] in der Version 9.2.4 verwendet.
Die Analyse der Mitschnitte der Kommunikation während der Benutzung der App[1] zeigt, dass neben dem erwarteten Verkehr für die SBB App auch noch einige Verbindungen zu weiteren Endpunkten aufgebaut werden (Abbildung 12.1).

[1] Suche nach Zugverbindungen über die Texteingabe, durchgeführt am 06.06.2019

Endpoints

IP	DNS Name	Reverse Lookup
212.47.171.92	sbb-ssl.wemfbox.ch	212.47.171.92
35.157.243.61	p1.sbbmobile.ch	ec2-35-157-243-61.eu-central-1.compute.amazonaws.com
18.194.125.217	p1.sbbmobile.ch	ec2-18-194-125-217.eu-central-1.compute.amazonaws.com
52.58.221.247	p1.sbbmobile.ch	ec2-52-58-221-247.eu-central-1.compute.amazonaws.com
52.49.204.15	logs1407.xiti.com	ec2-52-49-204-15.eu-west-1.compute.amazonaws.com
54.171.180.56	logs1407.xiti.com	ec2-54-171-180-56.eu-west-1.compute.amazonaws.com
54.76.54.11	logs1407.xiti.com	ec2-54-76-54-11.eu-west-1.compute.amazonaws.com
172.217.168.46	clients4.google.com	zrh04s14-in-f14.1e100.net
34.246.67.219	api-prod.axonvibelabs.com	ec2-34-246-67-219.eu-west-1.compute.amazonaws.com

Abbildung 12.1: Kontaktierte Server während Nutzung der SBB Mobile App

Auffällig sind dabei die Verbindungen mit den DNS Namen *sbb-ssl.wemfbox.ch*, *logs1407.xiti.com* und *api-prod.axonvibelabs.com*.

12.1.1 NET-Metrix

Aufrufe wie beispielsweise *sbb-ssl.wemfbox.ch* werden von der Firma NET-Metrix genutzt [93], um für ihre Kunden Messungen von Kennzahlen wie die Anzahl Seitenaufrufe oder einzigartige Clients zu zählen [94].
Die SBB weist in ihren Hinweisen zum Datenschutz darauf hin, dass sie die Dienste der Firma NET-Metrix in Anspruch nehmen [95].

12.1.2 AT Internet

Die Verbindungen auf *logs1407.xiti.com* sind auf die Firma AT Internet zurückzuführen. AT Internet bietet Lösungen an, um Informationen über die Besucher von Webseiten zu sammeln und diese auszuwerten [96]. Auch in Falle von AT Internet ist ein entsprechender Hinweis in den Informationen zum Datenschutz bei der SBB zu finden. Zudem schreibt die SBB, dass die gesammelte IP-Adressen durch entfernen der letzten drei Stellen anonymisiert werden, bevor sie diese an AT Internet übermitteln [95].

12.1.3 Axon Vibe

Der Endpunkt hinter *api-prod.axonvibelabs.com* gehört zur Firma Axon Vibe. Axon Vibe ist ein Schweizer Unternehmen, welches im Bereich der smart mobility tätig ist [97]. Laut Medienbericht der SBB aus dem Jahre 2017 wurde das damals nur für iOS verfügbare Reise-Cockpit von Axon Vibe entwickelt [98]. In der aktuellen Version der SBB Mobile Preview App ist diese Funktion auch für Android verfügbar.
Während der analysierten Messung wurden das Reise-Cockpit zwar nicht genutzt, die Kommunikation könnte aber durch den Bedarf an Informationen über den Reiseverlauf erklärt werden.

12.2 20min Mobile App

Kennzahl	Wert (% von Total)
Total number of streams	524
Number of inadvisable cipher suite sused	1 (14%)
Amount of unsecured traffic (by # of packets)	10,984 (53.51%)
Traffic incoming (Bytes)	13,768.94 KB
Traffic incoming (Packets)	12,874
Traffic outgoing (Bytes)	1,497.40 KB
Traffic outgoing (Packets)	7,952

Tabelle 12.2: Kennzahlen der Auswertung für die 20min Mobile App

Für die Tests mit dem 20min Mobile App wurde die 20min App in der Version 9.4.8.4 verwendet.
Um die 20min App zu testen, wurden nacheinander mehrere Artikel geöffnet[2]. Dieses Vorgehen wurde etwas später ein zweites Mal wiederholt. Während den zwei Captures, welche je den Netzwerkverkehr von etwa einer Minute umfassen, wurden Verbindungen zu 110 verschiedenen externen IPs aufgebaut.

12.2.1 AppNexus

Eine grosse Anzahl der erkannten externen Endpunkte (11 von 110) sind auf Subdomains von adnxs.com zurückzuführen. Diese Domain gehört der Firma AppNexus, welche laut eigenen Angaben den weltweit grössten, unabhängigen Marktplatz für Online-Werbung unterhält [99]. Die Verbindungen mit diesen Domains sind für ca. 20% des gesamten gemessenen Verkehrs verantwortlich.
Die Datenschutzerklärung von *www.20min.ch* weist darauf hin, dass Daten über den Webseitenbesucher gesammelt und für die Darstellung von personalisierter Werbung genutzt werden, und verlinkt auf eine Liste mit Trackern [100], welche in der Tamedia-Gruppe zum Einsatz kommen. Die Firma AppNexus beziehungsweise deren Tracker wird laut dieser Liste für Werbung genutzt, was durch die Auswertung bestätigt werden kann.

12.2.2 Rubicon Project

Die Aufrufe der Subdomains von rubiconproject.com stechen nicht durch ihre Häufigkeit hervor, sondern durch die Tatsache, dass auf *token.rubiconproject.com*, neben *thumbnails.20min-tv.ch* (Abbildung 12.2), die einzige Verbindung aufgebaut wurde, welche mit einer laut IANA [30] nicht mehr als sicher erachteten Cipher Suite gesichert wurde.

[2] Durchgeführt am 08.06.2019

Target IP	DNS Name	Reverse Lookup	Nr.	Description	Recommended
69.173.144.149	token.rubiconproject.com	69.173.144.149	156	TLS_RSA_WITH_AES_128_GCM_SHA256	N
188.40.52.132	thumbnails.20min-tv.ch	thumbnails1.20min-tv.ch	156	TLS_RSA_WITH_AES_128_GCM_SHA256	N

Abbildung 12.2: Verbindungen mit nicht empfohlener Cipher Suite in der Mobile App von 20min

Rubincon Project ist eine Firma, welche einen Marktplatz für digitale Werbungen betreibt [101]. In den ausgewerteten Captures für die 20min Mobile App konnte nur eine einzige Verbindung auf eine Domain von Rubicon Project festgestellt werden. Die Liste der von der Tamedia-Gruppe verwendeten Tracker [100] beinhaltet keinen Hinweis auf die Nutzung von Rubicon Project für jedwede Zwecke.

12.2.3 Verhältnis zwischen Inhalt und zusätzlichen Daten

In den zwei analysierten Captures konnten nur 12 (11%) externe Endpunkte identifiziert werden, welche eindeutig zu 20min.ch gehören. Die restlichen 98 (89%) externen Endpunkte gehören zu Providern von Feeds, Werbung oder Analytics.

12.3 Speedtests

Im Rahmen der Analyse wurden Speedtests mit mobilen Apps von drei verschiedenen Anbietern durchgeführt. Die Auswertungen des aufgezeichneten Netzwerkverkehrs wurden zwischen den Anbietern verglichen.

12.3.1 Cnlab Speedtest

Kennzahl	**Wert (% von Total)**
Total number of streams	37
Number of inadvisable cipher suite sused	0 (0%)
Amount of unsecured traffic (by # of packets)	62,427 (99.94%)
Traffic incoming (Bytes)	31,742.13 KB
Traffic incoming (Packets)	33,751
Traffic outgoing (Bytes)	33,077.88 KB
Traffic outgoing (Packets)	28,678

Tabelle 12.3: Kennzahlen der Auswertung für die mobile App des cnlab Speedtest

Bei der Durchführung des Speedtest von cnlab [102] in der Version 2.7.2 wurden nur Verbindungen auf drei verschiedene externe Endpunkte beobachtet. 19 Verbindungen wurden zu einem Server der Swisscom aufgebaut, welcher beim Speedtest als Quelle und Ziel für die übermittelten Testdaten diente. Weiter wurden 2 Verbindungen zu einem Server der Firma cnlab und 1 Verbindung auf einen Server der Firma OpenSignal, welche Netzwerkmessungen durchführen und analysieren [103], aufgebaut. Mitarbeiter von cnlab versicherten, dass ihr Speedtest nichts mit der Firma OpenSignal zu tun hat. Ausserdem wiesen sie darauf hin, dass der Meteor Speedtest, welcher zum Zeitpunkt

der Messung auf dem Mobiltelefon installiert war, von OpenSignal entwickelt wurde. Unter diesen Umständen liegt die Vermutung nahe, dass Teile des Meteor Speedtest auch aktiv sind, wenn andere Applikationen verwendet werden.
Für den Speedtest von cnlab kann festgehalten werden, dass nur Verbindungen aufgebaut werden, welche für die Funktionalität wirklich benötigt werden.

12.3.2 Ookla Speedtest

Kennzahl	**Wert (% von Total)**
Total number of streams	164
Number of inadvisable cipher suite sused	3 (38%)
Amount of unsecured traffic (by # of packets)	100,529 (98.89%)
Traffic incoming (Bytes)	35,117.25 KB
Traffic incoming (Packets)	46,435
Traffic outgoing (Bytes)	63,926.49 KB
Traffic outgoing (Packets)	55,240

Tabelle 12.4: Kennzahlen der Auswertung für die mobile App des Ookla Speedtest

Bei der Auswertung des Verkehrs während einem Speedtest mit der mobilen Version des Ookla Speedtest [104] in der Version 4.4.9.54928 fällt auf, dass eine sehr grosse Zahl an Verbindungen aufgebaut wurden. Während der Messungen wurden Verbindungen zu 45 verschiedenen, externen Endpunkten aufgebaut.
Weiter ist auffallend, dass beinahe doppelt so viele Daten hoch- wie heruntergeladen werden. Dies ist für einen Speedtest ungewöhnlich, da Up- und Download meist mit der gleichen Datenmenge getestet werden. Die grossen Datenmengen in beide Richtungen sind auf einen einzelnen externen Endpunkt, einen Server von iWay, zurückzuführen (Abbildung 12.3).

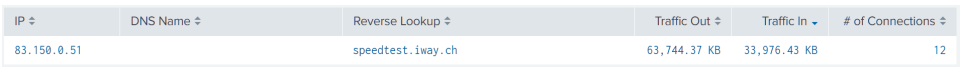


Abbildung 12.3: Datenverkehr des Ookla Speedtests mit einem Server von iWay

Neben den 12 Verbindungen, welche für die Durchführung des Speedtests benötigt wurden, gibt es eine Vielzahl an Verbindungen zu Ad-Exchanges (beispielsweise jenen von OpenX und Rubicon Project). Welchen Einfluss diese zusätzlichen Verbindungen auf die vom Speedtest ermittelten Werte für RTT, Up- und Downloadgeschwindigkeit haben, kann nicht beurteilt werden.

12.3.3 Meteor Speedtest

Kennzahl	Wert (% von Total)
Total number of streams	87
Number of inadvisable cipher suite sused	1 (25%)
Amount of unsecured traffic (by # of packets)	662 (0.60%)
Traffic incoming (Bytes)	56,364.95 KB
Traffic incoming (Packets)	57,879
Traffic outgoing (Bytes)	54,852.30 KB
Traffic outgoing (Packets)	52,944

Tabelle 12.5: Kennzahlen der Auswertung für die mobile App des Meteor Speedtest

Im Meteor Speedtest [105] in der Version 1.1.46 (107), welcher von OpenSignal entwickelt wurde, kann beobachtet werden, dass für den Up- und Download von Daten während einer Messung unterschiedliche externe Endpunkte verwendet werden (Abbildung 12.4).

IP	DNS Name	Reverse Lookup	Traffic Out	Traffic In	# of Connections
195.176.255.75	a195-176-255-75.deploy.akamaitechnologies.com	a195-176-255-75.deploy.akamaitechnologies.com	53,869.50 KB	1,184.44 KB	4
13.32.222.254	server-13-32-222-254.fra56.r.cloudfront.net	server-13-32-222-254.fra56.r.cloudfront.net	913.97 KB	55,045.46 KB	13

Abbildung 12.4: Datenverkehr des Meteor Speedtests mit unterschiedlichen Servern für Up- und Download

Dabei ist zudem interessant, dass es nicht nur verschiedene Server sind, sondern dass es sich dabei auch um unterschiedliche Provider handelt. Für die Messung der Downloadgeschwindigkeit wird ein Server von Amazon genutzt. Amazon AWS betreibt aktuell keine Server in der Schweiz [106]. Dieser spezifische Server steht in Frankfurt, Deutschland. Für die Messung der Uploadgeschwindigkeit wird hingegen ein Server von Akamai genutzt, welcher in der Schweiz steht. Auch wird der Datenverkehr für Up- und Download verschlüsselt übermittelt, was bei den zwei anderen getesteten Mobile Speedtests nicht beobachtet werden konnte.

In den Einstellungen der Meteor App findet sich eine Option, mit welcher definiert werden kann, wie oft die App im Hintergrund Speedtests durchführt. Diese Funktion ist standardmässig aktiviert und so konfiguriert, dass alle 5 Tage eine solche Messung durchgeführt wird. So wird die Applikation im Hintergrund dauerhaft zumindest teilweise aktiv sein, was die in Kapitel 12.3.1 gefundene Verbindung zu einem Server von OpenSignal erklären könnte.

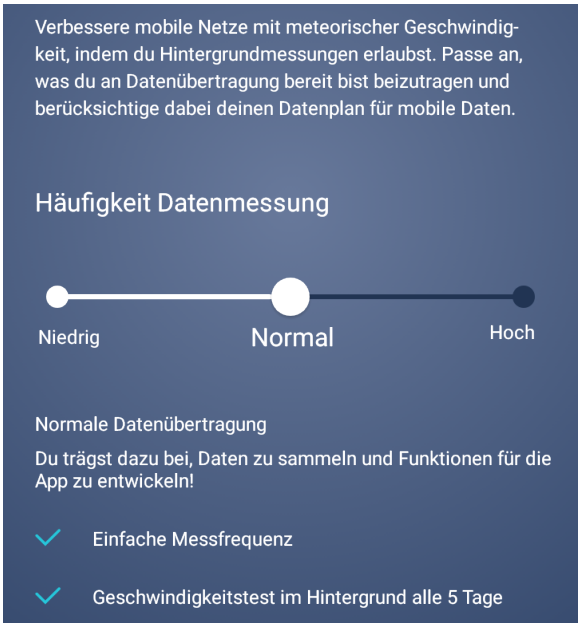


Abbildung 12.5: Einstellungen des Meteor Speedtests für die Häufigkeit der automatischen Datenmessungen

12.3.4 Vergleich der Speedtests

Die Analyse der Speedtest von cnlab, OpenSignal und Ookla mit dem **Traffic Analayzer** hat grosse Unterschiede aufgezeigt. So konnten sowohl bei der Applikation von OpenSignal, wie auch jener von Ookla eine grosse Anzahl an Verbindungen festgestellt werden, welche nicht mit der eigentlichen Messung der Geschwindigkeit in Verbindung stehen. Über den Zeitraum einer einzelnen Messung werden im Vergleich zur Applikation von cnlab beim Ookla Speedtest 127 Verbindungen zusätzlich aufgebaut, was einer Zunahme von ca. 343% entspricht. Auch die Applikation Meteor verzeichnet mit 50 Verbindungen mehr als jene von cnlab eine Zunahme von ca. 135%. Bei Ookla ist dies auf eine grosse Anzahl von Werbeprovidern zurückzuführen. Beim Meteor Speedtest auf die Nutzung diverser APIs.
Eine grössere Anzahl an Verbindungen geht jeweils auch mit einer grösseren Anzahl an genutzten Cipher Suites einher, welche nicht mehr verwendet werden sollten. Während beim Speedtest von cnlab keine unsicheren Cipher Suites erkannt wurden, waren es bei Meteor eine nicht mehr empfohlene Cipher Suite und bei Ookla drei.

12.4 Vergleich von Aufrufen mit und ohne Adblocker

Für den Vergleich zwischen dem Netzwerkverkehr mit und ohne Adblocker wurde auf dem Mobiltelefon das Plugin für den Adblocker uBlock in Version 1.19.6 im Browser Firefox installiert. Es wurde je ein Test mit und ein Test ohne Adblocker durchgeführt.

Bei beiden Tests wurde vorgängig der Cache des Browsers gelöscht. Zudem wurden dieselben 5 Artikel in beiden Test in der selben Reihenfolge aufgerufen.

12.4.1 Messung mit Adblocker

Kennzahl	Wert (% von Total)
Total number of streams	98
Number of inadvisable cipher suite sused	0 (0%)
Amount of unsecured traffic (by # of packets)	764 (2.64%)
Traffic incoming (Bytes)	34,553.32 KB
Traffic incoming (Packets)	25,126
Traffic outgoing (Bytes)	313.05 KB
Traffic outgoing (Packets)	3,758

Tabelle 12.6: Kennzahlen der Auswertung für die mobile Webseite von 20min (m.20min.ch) mit Adblocker

Der durchgeführte Test mit aktiviertem Adblocker weist keine unsicheren Cipher Suites auf. Der ausgehende Verkehr entspricht etwa 1% des einkommenden Verkehrs. Am stärksten fällt eine einzelne Verbindung zu einem Server von imgur.com ins Gewicht (ca. 29 MB im Download). Eine Suche nach der Ursache dafür in den besuchten Artikel auf 20min.ch hat ergeben, dass es sich dabei um eine einzelne GIF-Datei gehandelt hat [107].

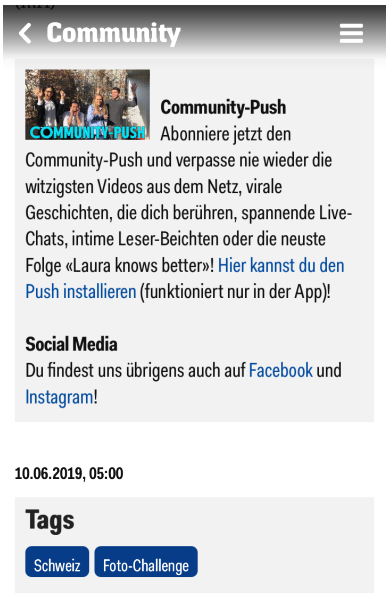


Abbildung 12.6: Screenshot des GIFs auf der mobilen Webseite von 20min.ch

Wird diese Datei nicht berücksichtigt, entspricht der eingehende Verkehr etwa 5% des

ausgehenden Verkehrs.

12.4.2 Messung ohne Adblocker

Kennzahl	Wert (% von Total)
Total number of streams	443
Number of inadvisable cipher suite sused	2 (25%)
Amount of unsecured traffic (by # of packets)	2,925 (6.13%)
Traffic incoming (Bytes)	48,512.98 KB
Traffic incoming (Packets)	38,081
Traffic outgoing (Bytes)	1,414.12 KB
Traffic outgoing (Packets)	9,610

Tabelle 12.7: Kennzahlen der Auswertung für die mobile Webseite von 20min (m.20min.ch) ohne Adblocker

Die Analyse der Messung ohne Adblocker zeigen, dass zwei unsichere Cipher Suites zum Einsatz kammen, was 25% aller verwendeten Cipher Suites ausmacht. Bei der Analyse der Datenmenge im Download zeigt sich dasselbe Verhalten wie in Kapitel 12.4.1 betreffend dem Laden des grossen GIFs. Mit dem GIF entspricht die Datenmenge im Upload ca. 2.9% des Downloads. Wird das GIF nicht beachtet steigt diese Zahl auf 6.7%.

12.4.3 Differenzen

Kennzahl	mit AB (%)	ohne AB (%)	Diff.
Total number of streams	98	443	+ 352%
Number of inadvisable cipher suite sused	0 (0%)	2 (25%)	NA
Amount of unsecured traffic (by # of packets)	764 (2.64%)	2,925 (6.13%)	+ 283%
Traffic incoming (Bytes)	34,553.32 KB	48,512.98 KB	+ 40%
Traffic incoming (Packets)	25,126	38,081	+ 52%
Traffic outgoing (Bytes)	313.05 KB	1,414.12 KB	+ 352%
Traffic outgoing (Packets)	3,758	9,610	+ 156%

Tabelle 12.8: Vergleich der Kennzahlen für die mobile Webseite von 20min (m.20min.ch) mit und ohne Adblocker (AB)

Die Werte in der Tabelle 12.8 zeigen eine Zunahme bei allen ausgewerteten Kennzahlen an, wenn dieselbe Messung wie zuvor mit aktiviertem Adblocker ohne Adblocker wiederholt wird.

12.5 Sonos mit Alexa

Kennzahl	Wert (% von Total)
Total number of streams	170
Number of inadvisable cipher suite sused	0 (0%)
Amount of unsecured traffic (by # of packets)	58,636 (95.86%)
Traffic incoming (Bytes)	530.24 KB
Traffic incoming (Packets)	2,418
Traffic outgoing (Bytes)	1,497.61 KB
Traffic outgoing (Packets)	5,180

Tabelle 12.9: Kennzahlen der Auswertung für Messungen über 14 Stunden mit der Sonos

Während einer 14-stündige Aufzeichnung des Verkehrs einer Sonos, ohne mit ihr in dieser Zeit zu interagieren, wurden nur 5180 Packets aus dem lokalen Netzwerk an einen externen Endpunkt verschickt. Der grösste Anteil an Packets, welche in der Tabelle 12.9 unter «Amount of unsecured traffic» zu sehen sind, wurde somit im lokalen Netzwerk versandt. Der grösste Anteil der Pakcets entfällt dabei auf das Spanning Tree Protocol (STP) (Abbildung 12.7).

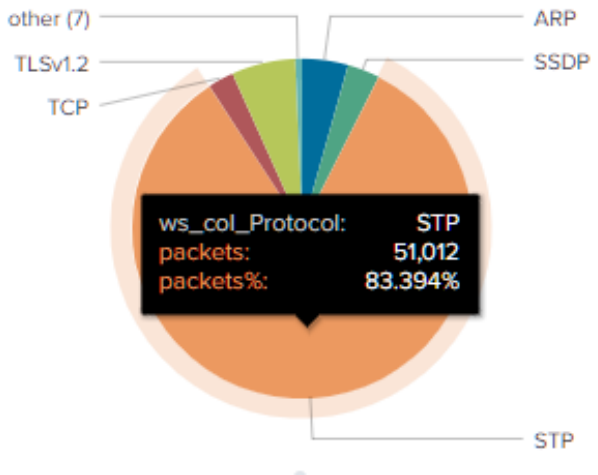


Abbildung 12.7: Anteil des STP Protokolls an gesamter Anzahl übermittelter Packets in Messung mit Sonos

Durch das Setup der Messung wäre es zu erwarten gewesen, an dieser Stelle keinen STP-Verkehr zu sehen. Das Auftreten von STP kann daher auf die Sonos als Verursacher zurückgeführt werden. Die Auswertung zeigt zudem auf, dass es sich nicht um eine zeitlich begrenzte Nutzung von STP gehandelt hat. Über den ganzen Verlauf der ausgewerteten Messung konnte eine konstante Nutzung des Protokolls gezeigt werden (Abbildung 12.8).

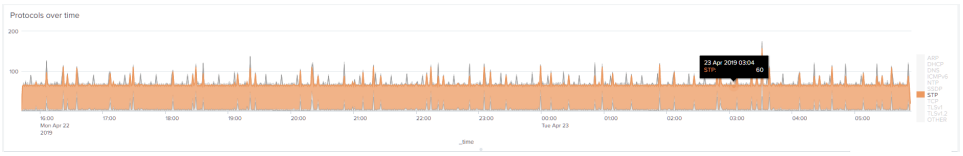


Abbildung 12.8: Nutzung des STP Protokolls über Zeit in Messung mit Sonos

Recherchen haben gezeigt das es sich dabei um ein bekanntes Verhalten handelt, welches jedoch gerade bei nichtsahnenden Benutzern Probleme verursachen kann. Die Firma Lode Audio schreibt beispielsweise, dass darauf geachtet werden soll, dass STP im Netzwerk korrekt konfiguriert wurde bevor eine Sonos zum Einsatz kommt. Sonst könne es vorkommen, dass diese zur Root Bridge auserkoren wird [108]. Sollte dies passieren kann die Performance des gesamten lokalen Netzwerks darunter leiden.

13 Schlusswort

Im Rahmen dieser Bachelorarbeit wurde eine Applikation zur Auswertung von Netzwerkcaptures auf Basis der Software «Splunk» entwickelt. Diese Applikation erlaubt es dem Benutzer sich einen Überblick über die Geschehnisse im Netzwerk zu machen. Unter anderem kann eingesehen werden, welches die aktivsten internen Systeme sind oder welche Endpunkte im Internet am meisten angesprochen wurden. Auch ist es damit möglich zu zeigen, ob veraltete Cipher Suites verwendet werden und welche Geräte diese nutzen.
Mit Netzwerkcaptures, welche erstellt wurden, um verschiedene Situationen analysieren zu können, konnte die Tauglichkeit der entwickelten Applikation für die vorgängig eruierten Anwendungsszenarien gezeigt werden. Zudem konnten mit den dafür durchgeführten Analysen gezeigt werden, dass die Nutzung eines Adblocker die Anzahl aufgebauter Verbindungen und die Menge an Daten, welche ins Internet übermittelt werden, signifikant einschränkt. Die Analysen brachten auch zu Tage, dass die Sonos das Spanning Tree Protocol (STP) Verkehr generiert, was je nach Konfiguration des eigenen Local Area Network (LAN) zu Problemen mit der Performance im Netzwerk führen kann. Ein Vergleich der Speedtests von cnlab, Ookla und OpenSignal zeigte Unterschiede in der Anzahl an aufgebauten Verbindungen und der Nutzung von externen Ressourcen und Trackern auf.

14 Ausblick

Während der Bachelorarbeit wurde die entwickelte Applikation nur auf Splunk innerhalb eines Docker Containers mit Debian genutzt. Sobald die Kompatibilität zu Windows gewährleistet werden kann, wird eine Veröffentlichung von **Traffic Analyzer** auf Splunkbase ins Auge gefasst.
Weiter soll die Applikation auf die Nutzung von Tshark Version 3.x umgestellt werden, sobald diese in Debian implementiert wurde. Dafür müssen nur die Regex-Regeln im Code entfernt werden, welche aktuell genutzt werden, um Befehle von der Syntax der Version 3.x auf solche der Version 2.x umzuschreiben.
In einer kommenden Arbeit könnte **Traffic Analyzer** um eine Funktion zur Zuordnung von Netzwerkverkehr auf die verursachende Applikation erweitert werden.

Abbildungsverzeichnis

Akronyme

API Application Programming Interface.

ARP Address Resolution Protocol.

BFS Bundesamt für Statistik.

CD Continuous Deployment.

CDN Content Delivery Network.

CI Continuous Integration.

CIDR Classless Inter-Domain Routing.

CLI Command Line Interface.

CSS Cascading Style Sheets.

CSV Comma Separated Values.

DDOS Distributed Denial of Service.

DHCP Dynamic Host Configuration Protocol.

DNS Domain Name System.

DNSSEC Domain Name System Security Extensions.

DoH DNS over HTTPS.

DoT DNS over TLS.

FURPS **F**unctionality, **U**sability, **R**eliability, **P**erformance / **P**ortability, **S**upportability.

GIF Graphics Interchange Format.

GPS Global Positioning System.

GUI Graphical User Interface.

HEX Hexadezimal.

HTTP Hypertext Transfer Protocol.

HTTPS Hypertext Transfer Protocol Secure.

IANA Internet Assigned Numbers Authority.

ID Identifier.

IDE Integrated Development Environment.

IEEE Institute of Electrical and Electronics Engineers.

IoT Internet of Things.

IP Internet Protocol.

IPv4 Internet Protocol Version 4.

IPv6 Internet Protocol Version 6.

ISP Internet Service Provider.

JSON JavaScript Object Notation.

LAN Local Area Network.

MAC Media Access Control.

MELANI Melde- und Analysestelle Informationssicherung.

NFR Non-Functional Requirement.

NTP Network Time Protocol.

OSI Open Systems Interconnection.

PCAP Packet Caputre.

PCAPNG Packet Caputre Next Generation.

PEP Python Enhancement Proposal.

rDNS Reverse DNS.

RFC Requests For Comment.

RTB Real Time Biding.

RTT Round Trip Time.

SCM Source Code Management.

SCSS Sassy CSS.

sed stream editor.

SIEM Security Information and Event Management.

SMTP Simple Mail Transfer Protocol.

SPL Search Processing Language.

SQL Structured Query Language.

SSID Service Set Identifier.

SSL Secure Socket Layer.

STP Spanning Tree Protocol.

TCP Transmission Control Protocol.

TLS Transport Layer Security.

TTL Time to live.

UDP User Datagram Protocol.

UI User Interface.

UX User Experience.

WLAN Wireless Local Area Network.

XML Extensible Markup Language.

Literatur

[1] *Internetnutzung mindestens 1 Mal pro Woche, internationaler Vergleich - 2005-2018 | Diagramm*, 2018. Adresse: https://www.bfs.admin.ch/bfs/de/home/statistiken/kultur-medien-informationsgesellschaft-sport/informationsgesellschaft/gesamtindikatoren/haushalte-bevoelkerung/internetnutzung.assetdetail.6786270.html (besucht am 2019-05-17).

[2] *Digitale Kompetenzen, Schutz der Privatsphäre und Online-Bildung: die Schweiz im internationalen Vergleich - Erhebung zur Internetnutzung 2017 | Publikation*, 2018. Adresse: https://www.bfs.admin.ch/bfsstatic/dam/assets/5306733/master (besucht am 2019-05-17).

[3] *Dauer der Internetnutzung pro Woche in der Schweiz - 2017 | Diagramm*, 2017. Adresse: https://www.bfs.admin.ch/bfs/de/home/statistiken/kultur-medien-informationsgesellschaft-sport/informationsgesellschaft/gesamtindikatoren/haushalte-bevoelkerung/internetnutzung.assetdetail.3862001.html (besucht am 2019-05-17).

[4] P. Heinzmann, *Zertifikatslehrgang*. Adresse: https://datenschutzkurs.ch/ (besucht am 2019-06-13).

[5] Craig Kanarick, Otto Timmons und Joe McCambley, *The 'First' Banner Ad*, 2014-10-28. Adresse: http://thefirstbannerad.com/ (besucht am 2019-05-19).

[6] A. LaFrance, *The First-Ever Banner Ad on the Web*, 2017. Adresse: https://www.theatlantic.com/technology/archive/2017/04/the-first-ever-banner-ad-on-the-web/523728/ (besucht am 2019-03-19).

[7] *Programmatic vs RTB. Differences Between Programmatic buying and RTB*. Adresse: https://smartyads.com/blog/stop-confusing-programmatic-with-real-time-bidding/ (besucht am 2019-06-07).

[8] *What does RTB Mean in Marketing? – SmartyAds*. Adresse: https://smartyads.com/blog/why-marketers-should-care-about-real-time-bidding/ (besucht am 2019-06-07).

[9] A. Hörler und J. Kugler, *Dual-WAN Router zur Verbesserung der User Experience*, 2018. Adresse: https://eprints.hsr.ch/747/ (besucht am 2019-06-12).

[10] *Zalando #whereveryouare Casefilm - YouTube*, 2016. Adresse: https://www.youtube.com/watch?v=Gw1pkngDq7w (besucht am 2019-03-15).

[11] J. Kastrenakes, *Adblock Plus now sells ads*, 2016. Adresse: https://www.theverge.com/2016/9/13/12890050/adblock-plus-now-sells-ads (besucht am 2019-06-07).

[12] *Google Chrome Built In Ad Blocker To Be Released July 2019*, 2019-01-17. Adresse: https://www.youtube.com/watch?v=AvvUnz1ZmLI (besucht am 2019-06-07).

[13] S. Shankland, *Developers are up in arms over a Google change that could cripple their Chrome ad-block extensions*. Adresse: https://www.cnet.com/news/google-may-break-ad-blockers-with-upcoming-chrome-change/ (besucht am 2019-06-07).

[14] J. Schallaböck, *Was ist und wie funktioniert Webtracking?* Adresse: https://irights.info/artikel/was-ist-und-wie-funktioniert-webtracking/23386.

[15] *Six Straight Years! Splunk Named a Leader in the Gartner SIEM Magic Quadrant*, 2019-05-04. Adresse: https://www.splunk.com/blog/2018/12/06/six-straight-years-splunk-named-a-leader-in-the-gartner-siem-magic-quadrant.html (besucht am 2019-05-07).

[16] *Parts of a Splunk Enterprise App | Splunk*. Adresse: http://dev.splunk.com/view/dev-guide/SP-CAAAE29 (besucht am 2019-05-06).

[17] *SIEM, AIOps, Application Management, Log Management, Machine Learning, and Compliance | Splunk*, 2019-04-30. Adresse: https://www.splunk.com/ (besucht am 2019-05-05).

[18] *SPL: Splexicon*. Adresse: https://docs.splunk.com/Splexicon:SPL (besucht am 2019-06-10).

[19] *About configuration files*. Adresse: https://docs.splunk.com/Documentation/Splunk/latest/Admin/Aboutconfigurationfiles (besucht am 2019-05-19).

[20] *props.conf - Splunk Documentation*. Adresse: https://docs.splunk.com/Documentation/Splunk/latest/Admin/Propsconf (besucht am 2019-05-19).

[21] *Stanza: Splexicon*. Adresse: https://docs.splunk.com/Splexicon:Stanza (besucht am 2019-05-19).

[22] *Create a Splunk app*. Adresse: http://dev.splunk.com/view/webframework-developapps/SP-CAAAEUC (besucht am 2019-05-06).

[23] „Automatic Generation of Mobile App Signatures from Traffic Observations“, Adresse: https://web.eecs.umich.edu/~zmao/Papers/infocom15-flowr.pdf (besucht am 2019-03-01).

[24] Vincent F. Taylor, Riccardo Spolaor, Mauro conti, Ivan Martinovic, *Robust Smartphone App Identification Via Encrypted Network Traffic Analysis*, 2017. Adresse: https://arxiv.org/abs/1704.06099 (besucht am 2019-03-01).

[25] Bundeskanzlei, *SR 235.1 Bundesgesetz vom 19. Juni 1992 über den Datenschutz (DSG)*, 09.06.2019. Adresse: https://www.admin.ch/opc/de/classified-compilation/19920153/index.html#a4 (besucht am 2019-06-13).

[26] Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter, *Internet- und E-Mail-Überwachung: Leitfaden Internet- und E-Mailüberwachung am Arbeitsplatz (Privatwirtschaft)*. Adresse: https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/arbeitsbereich/ueberwachung-am-arbeitsplatz/internet--und-e-mail-ueberwachung.html (besucht am 2019-06-14).

[27] *Address Allocation for Private Internets*, 19.05.2019. Adresse: https://tools.ietf.org/html/rfc1918 (besucht am 2019-05-19).

[28] *Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan*, 19.05.2019. Adresse: https://tools.ietf.org/html/rfc4632 (besucht am 2019-05-19).

[29] *IP Version 6 Addressing Architecture*, 26.05.2019. Adresse: https://tools.ietf.org/html/rfc4291 (besucht am 2019-06-03).

[30] IANA, *Transport Layer Security (TLS) Parameters*, 2005. Adresse: https://www.iana.org/assignments/tls-parameters/tls-parameters.xhtml#tls-parameters-4 (besucht am 2019-04-04).

[31] Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter, *Übermittlung ins Ausland: Staatenliste*. Adresse: https://www.edoeb.admin.ch/edoeb/de/home/datenschutz/handel-und-wirtschaft/uebermittlung-ins-ausland.html (besucht am 2019-06-14).

[32] *Dynamic Host Configuration Protocol*, 02.06.2019. Adresse: https://tools.ietf.org/html/rfc2131#section-3 (besucht am 2019-06-06).

[33] *Dynamic Host Configuration Protocol (DHCP) and Bootstrap Protocol (BOOTP) Parameters*, 06.05.2019. Adresse: https://www.iana.org/assignments/bootp-dhcp-parameters/bootp-dhcp-parameters.xhtml (besucht am 2019-06-06).

[34] *Cisco Nexus 7000 Series NX-OS Security Configuration Guide, Release 4.1 - Configuring DHCP Snooping [Cisco Nexus 7000 Series Switches]*, 01.01.2013. Adresse: https://www.cisco.com/c/en/us/td/docs/switches/datacenter/sw/4_1/nx-os/security/configuration/guide/sec_nx-os-cfg/sec_dhcpsnoop.html (besucht am 2019-06-06).

[35] P. Mockapetris und ISI, *Domain names - implementation and specification*, 04.06.2019. Adresse: https://tools.ietf.org/html/rfc1035#section-4.1.1 (besucht am 2019-06-12).

[36] *Domain Name System (DNS) Parameters*, 06.05.2019. Adresse: https://www.iana.org/assignments/dns-parameters/dns-parameters.xhtml (besucht am 2019-06-06).

[37] *DNS Best Practices, Network Protections, and Attack Identification*, 01.01.2017. Adresse: https://www.cisco.com/c/en/us/about/security-center/dns-best-practices.html (besucht am 2019-06-13).

[38] *Specification for DNS over Transport Layer Security (TLS)*, 02.06.2019. Adresse: https://tools.ietf.org/html/rfc7858 (besucht am 2019-06-07).

[39] *DNS Queries over HTTPS (DoH)*, 02.06.2019. Adresse: https://tools.ietf.org/html/rfc8484 (besucht am 2019-06-07).

[40] *How DNSSEC Works*. Adresse: https://www.cloudflare.com/dns/dnssec/how-dnssec-works/ (besucht am 2019-06-07).

[41] *Content Delivery Network Learning Center*. Adresse: https://www.akamai.com/us/en/cdn/ (besucht am 2019-06-06).

[42] KeyCDN, *IP Location Finder | Detailed geolocation data and RESTful API*. Adresse: https://tools.keycdn.com/geo (besucht am 2019-05-05).

[43] *IP Geolocation API and GeoIP Lookup Pricing*. Adresse: https://ipgeolocation.io/pricing (besucht am 2019-05-28).

[44] *Pricing*. Adresse: https://ipstack.com/product (besucht am 2019-06-07).

[45] *Free IP Geolocation Database Downloads | DB-IP Lite*. Adresse: https://db-ip.com/db/lite.php (besucht am 2019-05-18).

[46] *GeoIP2 City Accuracy*. Adresse: https://www.maxmind.com/en/geoip2-city-accuracy-comparison (besucht am 2019-06-12).

[47] *How accurate are your GeoIP2 and GeoIP Legacy databases? | MaxMind Support Center*. Adresse: https://support.maxmind.com/geoip-faq/geoip2-and-geoip-legacy-databases/how-accurate-are-your-geoip2-and-geoip-legacy-databases/ (besucht am 2019-05-17).

[48] *Kansas family sues mapping company for years of 'digital hell'*, 2016. Adresse: https://www.theguardian.com/technology/2016/aug/09/maxmind-mapping-lawsuit-kansas-farm-ip-address (besucht am 2019-05-19).

[49] K. Hill, *How Cartographers for the U.S. Military Inadvertently Created a House of Horrors in South Africa*, 2019-01-09. Adresse: https://gizmodo.com/how-cartographers-for-the-u-s-military-inadvertently-c-1830758394 (besucht am 2019-05-19).

[50] *Geolocation API*. Adresse: https://developer.mozilla.org/en-US/docs/Web/API/Geolocation_API (besucht am 2019-06-08).

[51] *Developer Guide | Geolocation API |*, 19.12.2018. Adresse: https://developers.google.com/maps/documentation/geolocation/intro (besucht am 2019-06-08).

[52] *IP Version 6 Addressing Architecture*, 26.05.2019. Adresse: https://tools.ietf.org/html/rfc4291#section-2.6 (besucht am 2019-06-03).

[53] *Reserved IPv6 Subnet Anycast Addresses*, 02.06.2019. Adresse: https://tools.ietf.org/html/rfc2526#section-2 (besucht am 2019-06-13).

[54] *Ping Test*. Adresse: https://tools.keycdn.com/ping (besucht am 2019-06-12).

[55] Paul & Will, *Ping time between Zurich and New York*, 12.06.2019. Adresse: https://wondernetwork.com/pings/Zurich/New%20York (besucht am 2019-06-12).

[56] *VirtualHost Examples: Running several name-based web sites on a single IP address*. 01.01.2019. Adresse: https://httpd.apache.org/docs/current/vhosts/examples.html (besucht am 2019-06-08).

[57] *File Descriptor Limits: Apache HTTP Server Version 2.4*, 01.01.2019. Adresse: https://httpd.apache.org/docs/2.4/vhosts/fd-limits.html (besucht am 2019-06-02).

[58] *Common DNS Operational and Configuration Errors*, 2019-05-05. Adresse: https://tools.ietf.org/html/rfc1912 (besucht am 2019-05-17).

[59] *The Transport Layer Security (TLS) Protocol Version 1.3*, 2019-04-01. Adresse: https://tools.ietf.org/html/rfc8446 (besucht am 2019-04-10).

[60] Bodo Möller und Google Security Team, Hrsg., *This POODLE bites: exploiting the SSL 3.0 fallback*, 2014-10-15. Adresse: https://security.googleblog.com/2014/10/this-poodle-bites-exploiting-ssl-30.html (besucht am 2019-05-18).

[61] *Firefox 34.0, See All New Features, Updates and Fixes*, 2014. Adresse: https://www.mozilla.org/en-US/firefox/34.0/releasenotes/ (besucht am 2019-05-19).

[62] *The Transport Layer Security (TLS) Protocol Version 1.2*, 02.06.2019. Adresse: https://tools.ietf.org/html/rfc5246#section-7.4 (besucht am 2019-06-06).

[63] *NGINX + HTTPS 101: The Basics & Getting Started*, 2016. Adresse: https://www.nginx.com/blog/nginx-https-101-ssl-basics-getting-started/ (besucht am 2019-06-13).

[64] John Kennedy und Michael Satran, *TLS Cipher Suites in Windows 10 v1809*. Adresse: https://docs.microsoft.com/en-us/windows/desktop/secauthn/tls-cipher-suites-in-windows-10-v1809 (besucht am 2019-06-13).

[65] *SSL/TLS Strong Encryption: How-To*, 01.01.2019. Adresse: https://httpd.apache.org/docs/current/ssl/ssl_howto.html (besucht am 2019-06-13).

[66] Melde- und Analysestelle Informationssicherung MELANI, *Social Engineering*. Adresse: https://www.melani.admin.ch/melani/de/home/themen/socialengineering.html (besucht am 2019-06-09).

[67] *Ransom.Wannacry*. Adresse: https://www.symantec.com/security-center/writeup/2017-051310-3522-99 (besucht am 2019-06-13).

[68] ——, *Verschlüsselungstrojaner*. Adresse: https://www.melani.admin.ch/melani/de/home/themen/Ransomware.html (besucht am 2019-06-13).

[69] *Welcome to The Public Listing For IEEE Standards Registration Authority*, 2019-05-16. Adresse: https://regauth.standards.ieee.org/standards-ra-web/pub/view.html (besucht am 2019-05-19).

[70] *ipstack*. Adresse: https://ipstack.com/ (besucht am 2019-06-12).

[71] *ipgeolocation*. Adresse: https://ipgeolocation.io (besucht am 2019-06-12).

[72] *WPO-Foundation/wptagent*, 2019-05-23. Adresse: https://github.com/WPO-Foundation/wptagent/blob/master/internal/optimization_checks.py (besucht am 2019-06-06).

[73] *StevenBlack/hosts*. Adresse: https://github.com/StevenBlack/hosts/blob/master/alternates/social/hosts (besucht am 2019-06-06).

[74] *Safe Browsing Lookup API (v4): Overview*, 28.07.2016. Adresse: https://developers.google.com/safe-browsing/v4/ (besucht am 2019-06-06).

[75] Yoyo.org, 1999. Adresse: https://pgl.yoyo.org/adservers/serverlist.php?hostformat=plain;showintro=0 (besucht am 2019-02-27).

[76] *Admin Manual: inputs.conf*. Adresse: https://docs.splunk.com/Documentation/Splunk/latest/Admin/Inputsconf (besucht am 2019-06-10).

[77] *Admin Manual: props.conf*. Adresse: https://docs.splunk.com/Documentation/Splunk/latest/Admin/Propsconf (besucht am 2019-06-10).

[78] *Setup.xml examples*. Adresse: http://dev.splunk.com/view/SP-CAAAE9B#creds (besucht am 2019-06-10).

[79] *SimpleXML: Splexicon*. Adresse: https://docs.splunk.com/Splexicon:SimpleXML (besucht am 2019-06-10).

[80] *Searches power dashboards and forms*. Adresse: https://docs.splunk.com/Documentation/Splunk/7.2.6/Viz/Savedsearches (besucht am 2019-06-10).

[81] *Documentation*, 06.06.2019. Adresse: https://sass-lang.com/documentation (besucht am 2019-06-10).

[82] *Bind data using tokens*. Adresse: http://dev.splunk.com/view/webframework-developapps/SP-CAAAEQB (besucht am 2019-06-10).

[83] *Window.sessionStorage*. Adresse: https://developer.mozilla.org/en-US/docs/Web/API/Window/sessionStorage (besucht am 2019-06-10).

[84] *Update Python 2.7 EOL date by hugovk: Pull Request #344*. Adresse: https://github.com/python/devguide/pull/344 (besucht am 2019-06-11).

[85] Ned Batchelder, *Coverage.py — Coverage.py 4.5.3 documentation*, 2019-03-10. Adresse: https://coverage.readthedocs.io/en/v4.5.x/ (besucht am 2019-05-14).

[86] *PEP 8 – Style Guide for Python Code*. Adresse: https://www.python.org/dev/peps/pep-0008/ (besucht am 2019-05-12).

[87] *wireshark*. Adresse: https://tracker.debian.org/pkg/wireshark (besucht am 2019-06-11).

[88] *Wireshark 3.0.0 Release Notes*, 01.06.2019. Adresse: https://www.wireshark.org/docs/relnotes/wireshark-3.0.0.html (besucht am 2019-06-11).

[89] *anjo-hsr/Traffic-Analyzer: app_deployer.sh*. Adresse: https://github.com/anjo-hsr/Traffic-Analyzer/blob/master/app_deployer.sh#L114 (besucht am 2019-06-11).

[90] *anjo-hsr/Traffic-Analyzer: enrich_csv.py*. Adresse: https://github.com/anjo-hsr/Traffic-Analyzer/blob/master/backend/bin/main/enrich_csv.py#L43 (besucht am 2019-06-11).

[91] *The MIT License*, 10.06.2019. Adresse: https://opensource.org/licenses/mit-license.php (besucht am 2019-06-11).

[92] *SBB Preview*. Adresse: https://www.sbb.ch/de/fahrplan/mobile-fahrplaene/sbb-mobile-preview.html (besucht am 2019-06-13).

[93] *Produktinfos*. Adresse: https://www.net-metrix.ch/produkte/net-metrix-mobile/produktinfos (besucht am 2019-06-07).

[94] *Messung*. Adresse: https://www.net-metrix.ch/service/reglement/messung (besucht am 2019-06-07).

[95] *Datenschutz*. Adresse: https://www.sbb.ch/de/meta/legallines/datenschutz.html (besucht am 2019-06-07).

[96] *AT Internets Analytics Suite, Web Analytics für alle*. Adresse: https://www.atinternet.com/de// (besucht am 2019-06-07).

[97] *What We Do*, 07.06.2019. Adresse: https://axonvibe.com/what-we-do (besucht am 2019-06-07).

[98] *App SBB Mobile: Eine halbe Million neue App-Kunden und neue Funktionen für die Preview Nutzer*. Adresse: https://company.sbb.ch/de/medien/medienstelle/medienmitteilungen/detail.html/2017/5/2305-2 (besucht am 2019-06-07).

[99] *About*, 04.06.2019. Adresse: https://www.appnexus.com/about (besucht am 2019-06-10).

[100] Simon Marquard Tamedia, *Tamedia_TrackingTools_DE*, 2019-04-09. Adresse: https://www.tamedia.ch/tl_files/content/Group/Datschutzerklaerung/Tamedia_TrackingTools_DE.pdf (besucht am 2019-06-10).

[101] *Rubicon Project Programmatic Advertising Exchange*. Adresse: https://rubiconproject.com/ (besucht am 2019-06-10).

[102] *Speedtest - information technology research*. Adresse: https://www.cnlab.ch/speedtest/ (besucht am 2019-06-13).

[103] *Mobile Analytics and Insights*. Adresse: https://www.opensignal.com/ (besucht am 2019-06-12).

[104] *Speedtest Apps*. Adresse: https://www.speedtest.net/apps (besucht am 2019-06-13).

[105] *Meteor*. Adresse: https://meteor.opensignal.com/ (besucht am 2019-06-13).

[106] *Globale AWS Infrastruktur & Availability Zones – AWS*, 07.06.2019. Adresse: https://aws.amazon.com/de/about-aws/global-infrastructure/ (besucht am 2019-06-12).

[107] *20min.ch Community-Push Gif*, 02.04.2019. Adresse: https://i.imgur.com/YM6km6i.gif (besucht am 2019-06-12).

[108] *Tips for installing Sonos*. Adresse: http://www.lodeaudio.com/downloads/Tips-For-Installing-Sonos.pdf (besucht am 2019-06-13).

Anhänge

In der abgegebenen ZIP-Datei finden sich folgende Dokumente:

- Bachelorarbeit.pdf
- Abstract.txt
- Aufgabenstellung_BA.pdf
- Einverständinserklärung.pdf
- Erklärung_zur_Urheberschaft.pdf
- Kontaktadressen.txt
- Persönliche_Berichte.pdf
- Vereinbarung_zur_Verwendung_und_Weiterentwicklung_der_Arbeit.pdf
- Sitzungsprotokolle.zip
- traffic-analyzer.tar.gz
- Usabilty_Test_Felix_Kugler.pdf
- Usabilty_Test_Tobias_Saladin.pdf

15 Projekt Management

15.1 Projektplan

Wir haben zu Beginn des Projekts einen Projektplan erstellt, welchen wir in folgende Phasen unterteilten:

- Projektstart (grün)
- Analyse und Design (blau)
- Realisierung (rot)
- Dokumentation und Finalisieren der Software (violett)
- Zeitpuffer / Projektabschluss (grau)

Der ursprüngliche Plan wurde im Verlauf des Projekts leicht angepasst, um neuen Gegebenheiten zu entsprechen. Vor allem wurden die Meilensteine entsprechend den Inputs des Korreferenten nach der Zwischenpräsentation in der neunten Woche der Arbeit angepasst. Zuletzt wurde der Zeitplan am 03.06.2019 mit genauen Zeitangaben betreffend der Abgabe verschiedener Dokumente ergänzt.

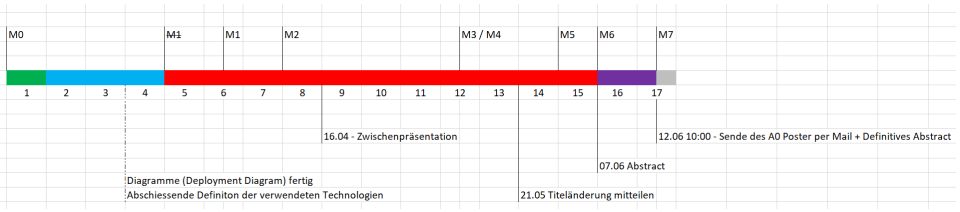


Abbildung 15.1: Projektplan Stand 03.06.2019

15.2 Arbeitszeiten

Die geleisteten Arbeitsstunden beliefen sich zum Schluss auf etwas mehr als die erwarteten 720 Stunden (360 Stunden pro Student). Dies lag vor allem daran, dass zum Schluss, als mit der entwickelten Applikation Messungen analysiert wurden, einige fehlende Funktionen und Bugs zum Vorschein kamen. Die Darstellung der geleisteten Stunden im Verlauf der Arbeit zeigt zudem auf, dass zu Beginn etwas gemächlich gestartet wurde. Die Schwankungen zwischen den Wochen und auch die Unterschiede in der Anzahl Stunden je Woche pro Student stammen grösstenteils von anderen Belastungen

im Studium und ausserhalb. Die Zunahme der Stunden in den letzten zwei Wochen war grundsätzlich geplant, da in diesen Wochen voll am Projekt gearbeitet werden konnte, fiel aber noch stärker aus als erwartet.

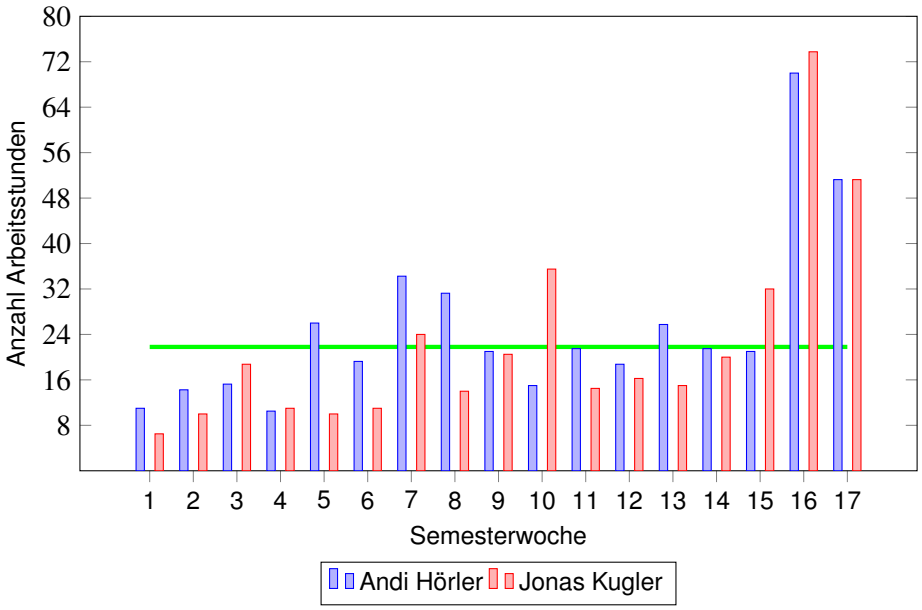


Abbildung 15.2: Arbeitszeiten pro Woche

Durch das Gruppieren der geleisteten Stunden nach der ausgeführten Tätigkeit zeigt sich, dass das Entwickeln der Applikation am meisten Zeit in Anspruch genommen hat. Der grosse Anteil an Stunden für die Analyse ist darauf zurückzuführen, dass darin sowohl der Aufwand für die Analyse der aktuellen Ausgangslage und der Anforderungen, wie auch jener für die Analyse der Auswertungen enthalten ist.

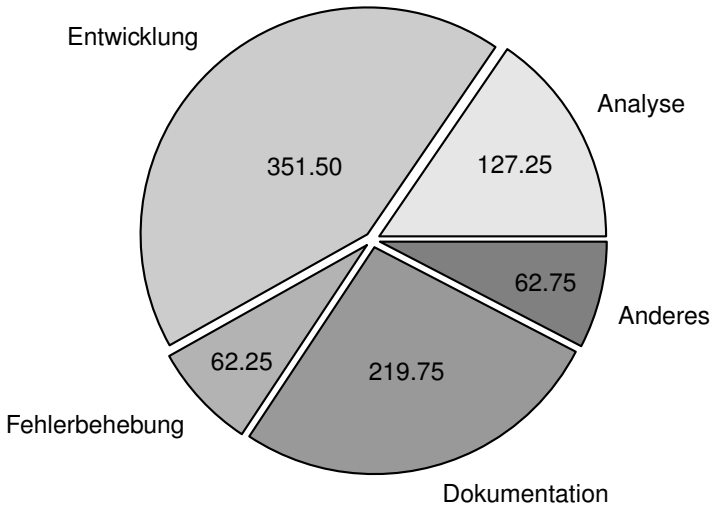


Abbildung 15.3: Arbeitszeiten nach Art des Aufwands