Zürcher Hochschule für Angewandte Wissenschaften

Department School of Engineering

Institut für angewandte Informationstechnologie

BACHELORARBEIT

Erstellung eines Digitalen Twins als Basis für die Cybersicherheit von Unternehmen

---

Autor:
Juvan Thavalingam
Tim Müller

Betreuer:
Prof. Dr. Ariane Trammell
Dr. Stephan Neuhaus

Eingereicht am
Juni 06, 2025

Studiengang:
Informatik, B.Sc.

Impressum

Projekt:	Bachelorarbeit
Titel:	Erstellung eines Digitalen Twins als Basis für die Cybersicherheit von Unternehmen
Autor:	Juvan Thavalingam Tim Müller
Datum:	Juni 06, 2025
Schlüsselwörter:	Cybersicherheit, Digital Twin, Data Analysis, Webpage, Database, Passiv OSINT
Copyright:	Zürcher Hochschule für Angewandte Wissenschaften

Studiengang:
Informatik, B.Sc.
Zürcher Hochschule für Angewandte Wissenschaften

Hauptbetreuerin:
Prof. Dr. Ariane Trammell
Zürcher Hochschule für Angewandte Wissenschaften
Email: ariane.trammell@zhaw.ch
Web: Link

Nebenbetreuer:
Dr. Stephan Neuhaus
Zürcher Hochschule für Angewandte Wissenschaften
Email: stephan.neuhaus@zhaw.ch
Web: Link

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig, oder gemeinsam mit den aufgeführten Gruppenmitgliedern, verfasst habe.

Ich habe ausschliesslich die im Text oder Anhang angegeben Quellen und Hilfsmittel (auch Internetseiten und generative KI-Tools) benutzt. Ich trage die Verantwortung für die Qualität des Textes sowie die Auswahl aller Inhalte und habe sichergestellt, dass Informationen und Argumente mit geeigneten wissenschaftlichen Quellen belegt bzw. gestützt werden. Generative KI-Tools wurden summarisch unter der Angabe des Namens und des Zwecks angegeben.

Bei Verfehlungen aller Art treten die Paragraphen 39 und 40 (Unredlichkeit und Verfahren bei Unredlichkeit) der ZHAW-Rahmenprüfungsordnung sowie die Bestimmungen der Disziplinarmassnahmen der Hochschulordnung in Kraft.

Ort, Datum:

Winterthur, 06. Juni 2025

Name:

Juvan Thavalingam

Ort, Datum:

Winterthur, 06. Juni 2025

Name:

Tim Müller

Abstract

In this bachelor's thesis, an application was developed that creates a digital twin as a basis for the cyber security of companies. A digital twin is a virtual image of an object or process, in this case a company. Such digital twins are usually used to analyse, improve or simulate processes.

The developed application performs passive OSINT and collects publicly accessible information about a company, including IPv4 and IPv6 addresses, mail servers, name servers, certificates, subdomains, e-mail addresses and telephone numbers and displays it clearly on a website. By focussing on Passive OSINT, all possibilities can be exploited, right up to technical and methodological limits.

The following technologies are used:

- **Flutter** used for the visual presentation and development of the user interface (frontend).
- **Phyton scripts** take over the collection of information in the backend.
- **PostgreSQL** serves as a relational database for storing the collected information.
- **Docker** enables the containerisation of the entire application and thus a portable, isolated infrastructure.

The architecture has a modular structure. A Python script is responsible for exactly one attribute (e.g. IPv4). In addition, a plugin system was implemented that enables an expandable backend structure. New attributes can thus be integrated clearly and easily, which significantly improves the maintainability and expandability of the application. This also makes it easier for security experts who do not have in-depth software development experience to get started. This gives them the opportunity to make changes to the code quickly and easily.

The application was able to collect a large amount of correct and relevant data, which was then used for security analyses. The information collected was verified and confirmed on site at two companies.

Despite the positive results, it became apparent that the application requires further development:

- Connection errors occasionally occur with some attributes, which can lead to no results being returned and the analysis having to be restarted.
- Analysing email addresses and telephone numbers is currently very time-consuming.

There is potential for optimization here, particularly with regard to performance and the completeness of the information recorded. Nevertheless, the results so far show that the approach developed works and is suitable for practical use. In future, the system should be supplemented with further attributes and improved in terms of user-friendliness and robustness.

Zusammenfassung

In dieser Bachelorarbeit wurde eine Applikation entwickelt, die einen Digital Twin als Grundlage für die Cybersicherheit von Unternehmen erstellt. Ein Digital Twin ist ein virtuelles Abbild eines Objekts oder Prozesses, in diesem Fall eines Unternehmens. Solche digitalen Zwillinge werden in der Regel eingesetzt, um Prozesse zu analysieren, zu verbessern oder zu simulieren.

Die entwickelte Applikation führt passiv OSINT durch und sammelt öffentlich zugängliche Informationen über ein Unternehmen, darunter IPv4- und IPv6-Adressen, Mailserver, Nameserver, Zertifikate, Subdomains, E-Mail-Adressen sowie Telefonnummern und stellt diese übersichtlich auf einer Webseite dar. Durch das Fokussieren auf Passiv OSINT können sämtliche Möglichkeiten ausgeschöpft werden, bis hin zu technischen und methodischen Grenzen.

Es werden folgende Technologien verwendet:

- **Flutter** wird für die visuelle Darstellung und Entwicklung der Benutzeroberfläche (Frontend) verwendet.
- **Phyton-Skripte** übernehmen im Backend das Sammeln der Informationen.
- **PostgreSQL** dient als relationale Datenbank zur Speicherung der gesammelten Informationen.
- **Docker** ermöglicht die Containerisierung der gesamten Applikation und damit eine portable, isolierte Infrastruktur.

Die Architektur wurde modular aufgebaut. Ein Python-Skript ist jeweils für genau ein Attribut (z.B. IPv4) zuständig. Zusätzlich wurde ein Pluginsystem implementiert, das eine erweiterbare Backend-Struktur ermöglicht. Neue Attribute können dadurch klar und einfach integriert werden, was die Wartbarkeit und Erweiterbarkeit der Applikation deutlich verbessert. Zusätzlich ermöglicht dies einen vereinfachten Einstieg für Sicherheitsexperten, die keine tiefgehende Softwareentwicklungserfahrung besitzen. Sie erhalten dadurch die Möglichkeit, schnell und unkompliziert Änderungen am Code vorzunehmen.

Die Anwendung konnte eine Vielzahl korrekter und relevanter Daten erfassen, die anschliessend für Sicherheitsanalysen genutzt wurden. Bei zwei Unternehmen konnten die gesammelten Informationen vor Ort verifiziert und bestätigt werden.

Trotz der positiven Ergebnisse zeigte sich, dass die Applikation weiterentwickelt werden muss:

- Bei manchen Attributen treten vereinzelt Verbindungsfehler auf, was dazu führen kann, dass keine Ergebnisse zurückgeliefert werden und die Analyse neu gestartet werden muss.
- Die Analyse der E-Mail-Adressen und Telefonnummern ist aktuell sehr zeitintensiv.

Hier besteht Optimierungspotenzial, insbesondere hinsichtlich der Performance und der Vollständigkeit der erfassten Informationen. Dennoch zeigen die bisherigen Resultate, dass der entwickelte Ansatz funktioniert und praxistauglich ist. In Zukunft sollte das System um weitere Attribute ergänzt und hinsichtlich Benutzerfreundlichkeit und Robustheit verbessert werden.

Danksagung

An dieser Stelle möchten wir uns bei allen Personen bedanken, die uns bei der Anfertigung unserer Bachelorarbeit unterstützt haben.

Ein besonderer Dank gilt unseren beiden Betreuern Prof. Dr. Ariane Trammell und Dr. Stephan Neuhaus von der Forschungsgruppe Information Security der ZHAW. Sie haben uns nicht nur in unseren wöchentlichen Meetings mit wertvollem Feedback und hilfreichen Tipps fachlich begleitet, sondern auch die Nutzung der DeHashed- und Shodan-APIs durch finanzielle Unterstützung ermöglicht.

Zudem möchten wir uns bei den beiden Unternehmen bedanken, die sich die Zeit genommen haben, mit uns die Ergebnisse des Digital Twins durchzugehen und uns direktes, konstruktives Feedback zu geben.

Ohne diese Unterstützung wäre es nicht möglich gewesen, diese Bachelorarbeit erfolgreich abzuschliessen.

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

A-Record	**A**ddress Record für IPv4
AAAA-Record	**A**ddress Record für IPv6
API	**A**pplication **P**rogramming **I**nterface
CPE	**C**ommon **P**latform **E**numeration
CSS	**C**ascading **S**tyle **S**heets
CSRF	**C**ross-**S**ite **R**equest **F**orgery
CYREN	**C**y**R**e**N**
CVE	**C**ommon **V**ulnerabilities and **E**xposures
DDoS	**D**istributed-**D**enial-**o**f-**S**ervice-Angriff
DNS	**D**omain **N**ame **S**ystem
HTML	**H**yper**t**ext **M**arkup **L**anguage
HTTP	**H**yper**t**ext **T**ransfer **P**rotocol
HTTPS	**H**yper**t**ext **T**ransfer **P**rotocol **S**ecure
IANA	**I**nternet **A**ssigned **N**umbers **A**uthority
IP	**I**nternet**p**rotokoll
KMU	**K**lein und **M**ittel **U**nternehemen
MX-Record	**M**ail E**x**change Record
NS-Record	**N**ame **S**erver Record
OSINT	**O**pen **S**ource **I**ntelligent
PTR-Record	**P**oin**t**e**r** Record
SOA-Record	**S**tart **o**f **A**uthority Record
SQL	**S**tructured **Q**uery **L**anguage
TXT-Record	**Tex**t Record
TCP	**T**ransmission **C**ontrol **P**rotocol
URL	**U**niform **R**esource **L**ocator
XSS	C**(=X)**ross-**S**ite **S**cripting

Kapitel 1

Einleitung

Der Wohlstand der Schweiz basiert zu einem grossen Teil auf dem Erfolg der kleinen und mittleren Unternehmen (KMU), die rund 99% aller Schweizer Firmen ausmachen. [1] Gleichzeitig ist der technologische Fortschritt so weit fortgeschritten wie nie zuvor. Seit dem Aufkommen des Internets und der Verbreitung von Smartphones hat nahezu jeder Mensch Zugang zur digitalen Welt.

Im Jahr 2023 erfasste das Bundesamt für Statistik 43'389 digitale Straftaten. [2] Bereits ein Jahr später, 2024, stieg diese Zahl auf 59'034 Fälle. [3] Das ist ein Anstieg um rund 36%. Besonders betroffen sind dabei kritische Infrastrukturen wie Spitäler sowie KMUs. Hackerangriffe verursachen Schäden in Millionenhöhe und schwächen damit direkt die Wirtschaft.[4]

Gerade deshalb ist es entscheidend, die eigene digitale Infrastruktur zu schützen und nur die notwendigen Informationen öffentlich preiszugeben. In der Praxis ist es jedoch oft schwierig festzustellen, welche Daten und Informationen tatsächlich im Internet verfügbar sind. Solche öffentlich zugänglichen Daten hinterlassen digitale Spuren, welche, sofern ungeschützt, potenzielle Schwachstellen darstellen und Angriffsflächen für Cyberkriminelle bieten.

An dieser Stelle kommt der Begriff des digitalen Zwillings (Digital Twin) ins Spiel. Ein digitaler Zwilling ist ein digitales Abbild oder Modell eines bestehenden Projekts, sei es physischer oder virtueller Natur. Ziel ist es, möglichst viele Informationen über das Originalprojekt zu erfassen und diese präzise abzubilden. Beispiele hierfür sind Produktzwillinge (Fahrzeuge), Maschinenzwillinge (Maschinen und Anlagen) oder Prozesszwillinge. [5] Durch diese Modelle lassen sich systematisch Analysen durchführen und Optimierungspotenziale identifizieren.

In dieser Bachelorarbeit wurde der Versuch unternommen, einen digitalen Zwilling einer Firma zu erstellen. Hierzu wurden öffentliche Informationen wie IPv4- und IPv6-Adressen, Mailserver, Nameserver, Zertifikate, Subdomains, E-Mails und Telefonnummern gesammelt und in strukturierter Form auf einer Webseite visualisiert. So lassen sich die Daten gezielt kontrollieren und überprüfen, insbesondere daraufhin, ob sensible Informationen öffentlich zugänglich sind, die es nicht sein sollten. Bei Bedarf können so Sicherheitsmassnahmen ergriffen werden.

1.1 Ausgangslage

Im ersten Kapitel erfolgt die Beschreibung der Ausgangslage des Projektes. Zunächst werden bestehende Arbeiten dargestellt, die für die vorliegende Bachelorarbeit von Relevanz sind. Im weiteren Verlauf wird der Stand der Technik hinsichtlich der Erstellung eines Digital Twins erörtert. Abschliessend wird das Projekt Cyren des Kantons Zürich beleuchtet.

1.1.1 Bestehende Arbeiten

In diesem Abschnitt werden zwei relevante Arbeiten beschrieben, die für die vorliegende Bachelorarbeit von Bedeutung sind. Die erste Arbeit bildet die theoretische Grundlage für diese Bachelorarbeit und die zweite Arbeit befasst sich mit der Technik des Web-Scraping.

1.1.1.1 Digital Twin for Cybersecurity: Towards Enhancing Cyber Resilience

Im Jahr 2022 wurde die Arbeit „Digital Twin for Cybersecurity: Towards Enhancing Cyber Resilience“ von den Autoren Rajiv Faleiro, Lei Pan, Shiva Raj Pokhrel und Robin Doss veröffentlicht.
In dieser Untersuchung wurde die potenzielle Anwendung digitaler Twins zur Steigerung der Cyber-Resilienz von Unternehmen analysiert. Die Autoren identifizierten eine dringende Notwendigkeit für innovative Schutzmassnahmen angesichts der zunehmenden Bedrohungen für die Cybersicherheit von Unternehmen. Im Rahmen der vorliegenden Arbeit wurde eine systematische Literaturrecherche durchgeführt und ein Forschungsrahmen für digitale Twins in der Cybersicherheit entwickelt. Die Autoren dieser Arbeit gelangten zu dem Schluss, dass digitale Twins als Sicherheitsmechanismen zur Bedrohungserkennung und Prävention genutzt werden können. Als Herausforderung wurde von ihnen die Lücke zwischen dem physischen und dem digitalen Twin identifiziert. [6]
Die vorliegende Arbeit bestätigt, dass sich die Nutzung eines Digital Twins zur Verbesserung der Cybersecurity-Sicherheit eines Unternehmens als sinnvoll erweist. Ihre Arbeit ist daher für diese Bachelorarbeit von Relevanz.

1.1.1.2 Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application

Im Jahr 2021 wurde die vorliegende Arbeit von Moaiad Ahmad Khder im International Journal of Advanced Soft Computing and Applications (IJASCA) publiziert.

Die Untersuchung von Web Scraping und Web Crawling als Methoden zur automatischen Extraktion von Daten aus Webseiten sowie die Analyse der Techniken, Ansätze und Anwendungen dieser Methoden in verschiedenen Bereichen stellen die zentralen Anliegen der Arbeit dar. Zur Erreichung des Hauptziels bedienen sich die Autoren unterschiedlicher Methoden. Einerseits wurde eine umfassende Literaturrecherche über Web Scraping, Crawling und deren Anwendungen durchgeführt, andererseits erfolgte ein Vergleich verschiedener Scraping-Techniken wie HTML-Parsing, DOM-Manipulation und API-Zugriffe. Die Autoren kommen zu dem Schluss, dass Web Scraping eine leistungsstarke Methode ist, um grosse unstrukturierte Daten in eine strukturierte Form zu überführen. Die Analyse ergibt, dass HTML-Parsing zwar schnell, aber limitiert, DOM-Manipulation zwar effizient, aber aufwendig und API-Zugriff optimal, aber oft eingeschränkt ist. Als Herausforderungen werden in der Arbeit die technischen Barrieren wie Captchas, IP-Sperren

und dynamische Inhalte identifiziert.
Für die vorliegende Arbeit, die einen Digital Twin erstellen soll, werden Web-Scraping-Techniken genutzt. Aufgrund der Vergleichbarkeit der drei Techniken HTML-Parsing, DOM-Manipulation und API-Zugriff werden in dieser Arbeit zuerst die API-Zugriffe genutzt, da diese, wenn nicht möglich, durch eine möglichst schnelle Technik ersetzt werden sollte. [7]

1.1.2 Stand der Technik

Im Folgenden soll der gegenwärtige Stand der Technik bei der Erstellung eines Digital Twins einer Firma dargelegt werden. In diesem Zusammenhang werden sowohl Open-Source-Reconnaissance-Tools als auch Vulnerability Scanner erörtert.

1.1.2.1 Open-Source-Reconnaissance-Tool

Durch den Einsatz von Reconnaissance-Tools können Web-Aufklärungsaufgaben automatisiert werden, wodurch folgende Vorteile erzielt werden:
Die Ausführung wiederholender Aufgaben erfolgt schneller. Die Abdeckung einer grossen Anzahl von Domains wird erleichtert. Vordefinierte Einstellungen minimieren das Risiko menschlicher Fehler.

Ein Python-basiertes Aufklärungstool ist FinalRecon, das eine automatische Suche nach verschiedenen Informationen beinhaltet. Diese enthält folgende Bausteine:
Ein Whois-Lookup dient der Aufdeckung von Registrierungsdetails für Domains, während die SSL Certificate Information die Gültigkeit der Zertifikate sowie deren Stellen und weitere relevante Details aufzeigt. Die Header Information der Domain zeigt die verwendeten Serverdetails sowie Technologien. Ein Crawler sucht unter anderem nach Links in Javascripts sowie nach robots.txt und sitemap.xml. Darüber hinaus führt er automatisch eine DNS, Subdomain sowie Directory Enumeration durch. Abschliessend werden die URLs der letzten fünf Jahre in einer Web-Archive-Software aufgerufen, um eine Analyse der Webseitenänderungen sowie potenzieller Schwachstellen zu ermöglichen. [8]

Die nachfolgende Abbildung veranschaulicht eine Konsolenausgabe von FinalRecon. Deutlich erkennbar ist ein Apache-Server in der Version 2.4.41.

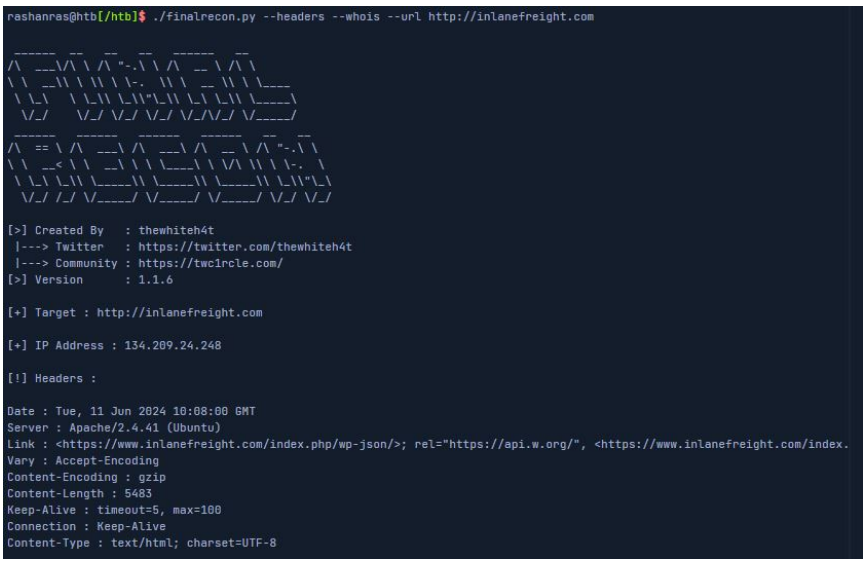

ABBILDUNG 1.1: Nessus Essentials [8].

1.1.2.2 Vulnerability Scanner

Ein Vulnerability Scanner ist ein Instrument, das darauf abzielt, automatisiert Schwachstellen in einem System zu identifizieren. Es nutzt definierte Angriffsmuster, um Schwachstellen zu lokalisieren und kann SQL-Injections testen. Darüber hinaus analysiert es die verwendeten Technologien, um zu ermitteln, ob bekannte Schwachstellen vorliegen. Die gefundenen Schwachstellen werden in einer Übersicht dargestellt, die auf dem Standard Common Vulnerabilities and Exposures (CVE) basiert. Zu diesem Zweck werden verschiedene Tools wie Nessus Essentials und OpenVAS eingesetzt. [9]

In der vorliegenden Bachelorarbeit wird auch nach CVE in der Infrastruktur von Unternehmen gesucht.

In der Abbildung ist ein Ausschnitt von Nessus Essentials dargestellt.

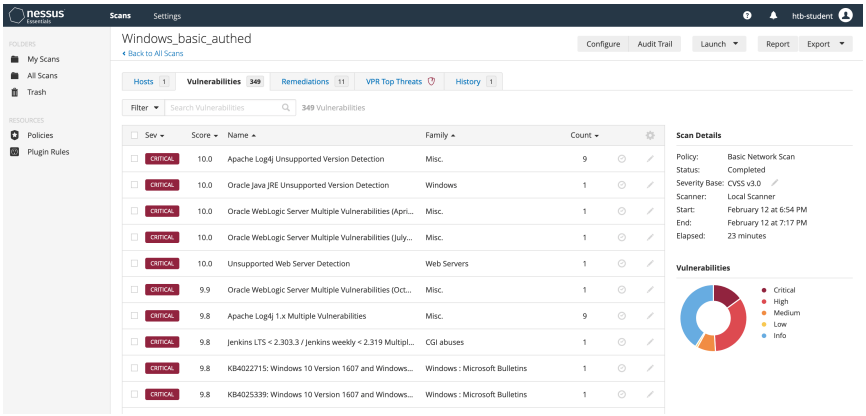

ABBILDUNG 1.2: Nessus Essentials [9].

1.1.3 Project Cyren ZH

Das Projekt „Cyber Resilience Network (CYREN)“ für den Kanton Zürich stellt eine Initiative zur Stärkung der Cyber-Abwehr des Kantons dar, die auf die zunehmenden Fälle von Cyberkriminalität reagiert.
Ziel des Projekts ist es, die Cyber-Sicherheitslage des Kantons durch verschiedene Massnahmen zu verbessern. Dazu zählen die Förderung von Forschung und Bildung im Bereich der Cybersicherheit, die Etablierung eines Netzwerkes zwischen Wissenschaft, Wirtschaft und öffentlichem Sektor sowie die praktische Unterstützung von Unternehmen und Behörden. [10]

Die vorliegende Bachelorarbeit ist Bestandteil eines Teilprojekts des Cyren.

1.2 Anforderungen & Motivation

Aus Sicht der zukünftigen Nutzenden der Webseite sind ein kostenloser Zugang zu einem Attack Surface Scan sowie eine einfache Bedienung von besonderer Relevanz. Somit können die Nutzenden unkompliziert ermitteln, welche Daten ihres Unternehmens bereits öffentlich im Internet zugänglich sind.
Auf der Webseite sollten die Daten sortiert und zusammengefasst werden, damit die Übersicht erhalten bleibt und die Firmen ihren Lehren daraus ziehen können. So ist einfach und schnell ersichtlich, welche Informationen im Netz zu finden sind, die eigentlich nicht öffentlich sein sollten (Passwörter, E-mails usw.).
Dieses Projekt wird in Zukunft in Projekt Cyren eingegliedert und hat daher spezielle Anforderungen.
Ein wesentlicher Aspekt ist dabei die Befolgung der geltenden Gesetzgebung. Zu diesem Zweck werden alle relevanten Informationen aus öffentlichen Quellen gesammelt und auf diese Weise legal auf passive OSINT-Technik zurückgegriffen. Passives OSINT ist ein Verfahren zur Sammlung von bereits öffentlich zugänglichen Informationen, die von Drittanbietern oder im Internet bereitgestellt werden. Diese Methode ist legal, da sie ausschliesslich auf bestehende, frei verfügbare Datenquellen zugreift. [11] Im Gegensatz dazu steht aktives OSINT, bei dem Informationen direkt beim Zielsystem erhoben werden. Dies ist ohne ausdrückliche Zustimmung des Betroffenen illegal und hat strafrechtliche Konsequenzen. In dieser Arbeit wird ausschliesslich auf passives OSINT fokussiert. Der Grund dafür liegt darin, zu versuchen, alle Möglichkeit von passiv OISNT auszureizen und diese gut umzusetzen.
Darüber hinaus ist eine leichte Erweiterbarkeit bzw. ein modularer Aufbau der Webseite ein wesentlicher Aspekt. Ein modularer Aufbau hat den Vorteil, dass der Applikation einfach und schnell neue Attribute hinzugefügt werden können. Dadurch wird die Weiterentwicklung und Wartung der Applikation erleichtert sowie die Einarbeitung neuer Entwickelnden und Sicherheitsexperten, die kaum Erfahrung in der Softwareentwicklung haben, unterstützt.

1.3 Zielsetzung

Dies ist eine Auflistung von Anforderung an unser Projekt:

- Die Webseite soll einen modularen Aufbau haben.
- Weitere Attribute sollen sich einfach ergänzen lassen.
- Die Webseite soll ansprechend und übersichtlich sein und um nützliche Features ergänzt werden.
- Der Scanvorgang soll OSINT-konform sein.

Kapitel 2

Architektur

In diesem Kapitel wird die Architektur der Applikation im Detail erläutert. Ziel war es, eine einfache, verständliche und klar strukturierte Architektur zu entwerfen, die bereits auf den ersten Blick nachvollziehbar ist.

Eine übersichtliche Architektur verbessert nicht nur die Lesbarkeit und Nachvollziehbarkeit des Projekts, sondern ermöglicht auch eine klare Trennung einzelner Komponenten in logische Ebenen.

Im Kapitel Abstrakte Architektur wird die vereinfachte Architektur der Applikation dargestellt. Hier wird bewusst mit einfachen Begriffen und grafischen Darstellungen gearbeitet, um einen schnellen und intuitiven Überblick zu vermitteln.

Für ein tiefergehendes technisches Verständnis wird das Kapitel Klassendiagramm empfohlen. Dort werden die einzelnen Klassen mit ihren Funktionen verbildlicht und deren Zweck sowie Zusammenspiel erläutert.

2.1 Abstrakte Architektur

Um einen groben Überblick zu erhalten, erläutert der folgende Text die vereinfachte Architektur. Die Applikation kann in zwei Ebenen unterteilt werden: Frontend und Backend. Frontend ist die Ebene, welche der Nutzende auf seinem Bildschirm sieht und Backend ist alles, was im Hintergrund abläuft. Im Backend läuft die ganze technische Logik und im Frontend wird die Benutzenden-Eingabe wahrgenommen.

Im Loginscreen kann sich der Nutzende einloggen und landet dann im Homescreen. Hier kann er einen Domainnamen einer Firma (z.B. zhaw.ch) eingeben und die Attribute auswählen (Zertifikate, E-Mail, Telefonnummer...), die ihn interessieren. Wenn der Nutzende alles ausgewählt hat, klickt er auf *scan*. Das löst einen Signal im Backend aus und via Plugin wird im Internet nach Informationen zu den ausgewählten Attributen gesucht und die Ergebnisse werden in einer Datenbank automatisch abgespeichert. Der Nutzende kann sich mit *GET* die Ergebnisse auf dem Bildschirm anzeigen lassen. Mit *export* kann die Person die Ergebnisse lokal herunterladen.

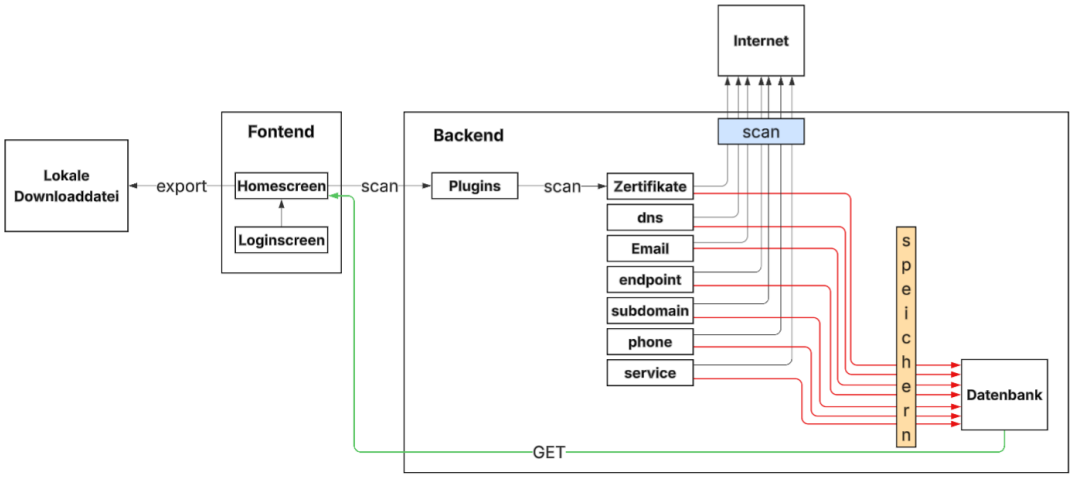

ABBILDUNG 2.1: Vereinfachte Architektur

2.2 Klassendiagramm

Das ist die Architekur der Applikation. Die Klasse login_screen ist das Anmelde- und Willkommensfenster. Das ist das erste Fenster, welches der Nutzende beim Besuch der Webseite sehen wird. Die Klasse home_screen ist das Hauptfenster. Hier können die ganzen Analysen gemacht und gespeichert werden.

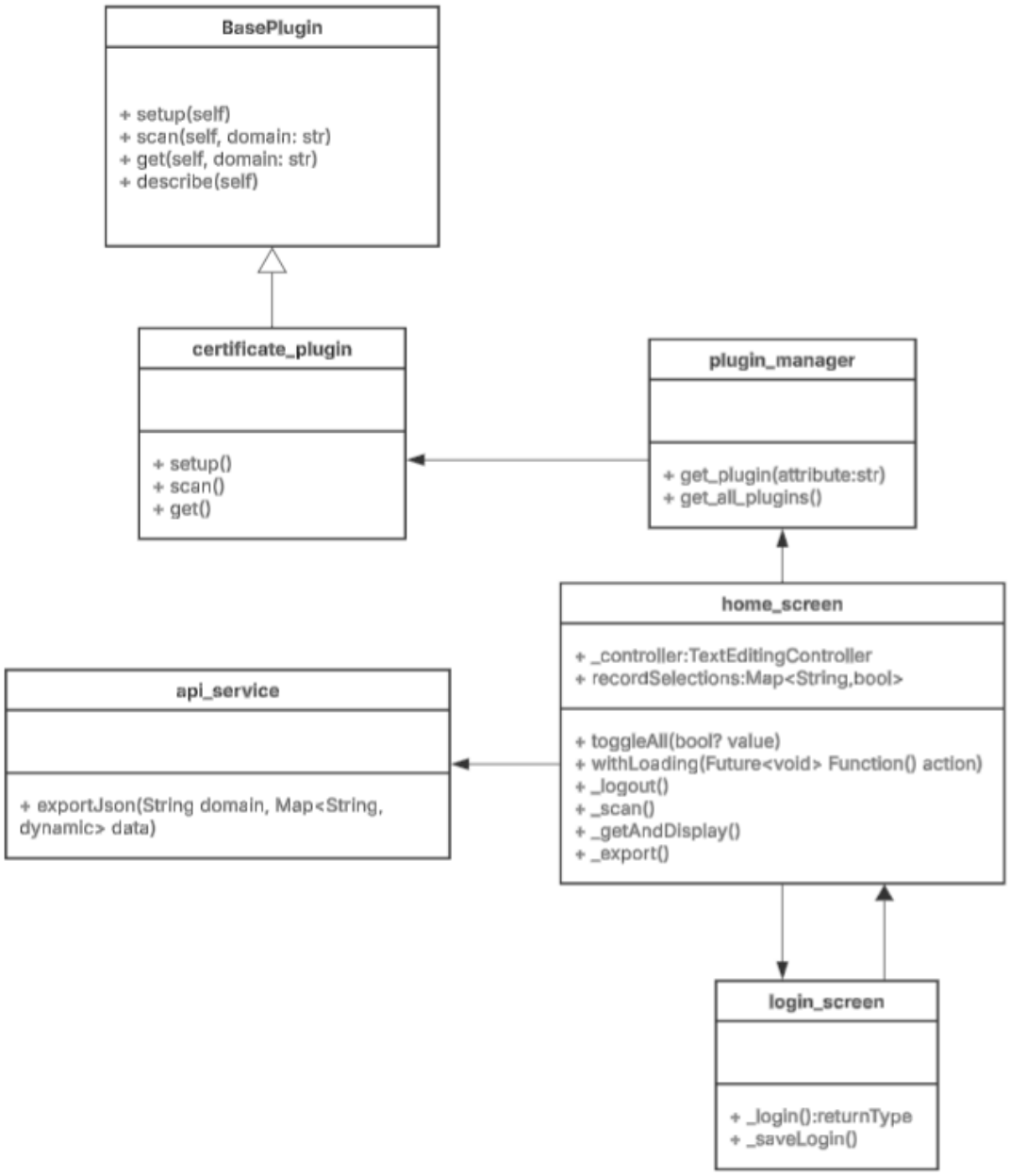

ABBILDUNG 2.2: Architektur der Applikation

Die Methoden und die Klassen im Klassendiagramm werden in den nachfolgenden Unterkapiteln detaillierter beschrieben.

2.2.1 login_screen

In dieser Klasse wurde das Login implementiert. Wenn man auf die Webseite geht, wird diese Klasse als Erstes angezeigt. Diese Klasse hat die zwei Methoden *_login()* und *_saveLogin()*.

2.2.1.1 _login()

Die Anmeldedaten werden überprüft und bei korrekter Eingabe wird der Nutzende zu *home_screen* weitergeleitet. Bei nicht korrekten Daten wird eine Fehlermeldung ausgegeben.

2.2.1.2 _saveLogin()

Der Loginstatus wird lokal abgespeichert. Das verbessert das Erlebnis für den Nutzenden, da dieser sich nicht die ganze Zeit anmelden muss.

2.2.2 home_screen

Dies ist das Hauptprogramm bzw. mit dieser Klasse können Firmen analysiert werden. Dabei existieren folgende Methoden:

- toggleAll(bool? value):
- withLoading(Future<void> Function() action):
- _logout()
- _scan()
- _getAndDisplay()
- _export()

2.2.2.1 toggleAll()

Auf einen Schlag werden alle Attribute ausgewählt.

2.2.2.2 WithLoading()

Diese Methode funktioniert wie ein Ladebildschirm und zeigt an, ob die Analyse der Attribute abschlossen ist oder ob diese noch weiter andauert. Wenn der Status von isLoading *true* ist, dann ist die Analyse noch nicht beendet und auf *false*, wenn sie abgeschlossen ist.

2.2.2.3 _logout()

Dies ist das Gegenstück von *_login()*. Mit *_logout()* kann man sich abmelden und zurück zur Login-Seite wechseln.

2.2.2.4 _scan()

Es werden alle ausgewählten Attribute analysiert und das Ergebnis wird in der Datenbank gespeichert. Das ist der Ablauf:

1. WithLoading() wird ausgeführt.
2. Aus dem Textfeld _controller wird die Domain gelesen und getrimmt.
3. Eine neue Map wird vorbereitet (statusMap), damit die Ergebnisse hier gespeichert werden können.
4. Für jedes ausgewählte Attribut wir das im Ordner *Backend* passende Plugin geladen.
5. Die Scan Methode des Plugins wird aufgerufen.
6. Das Ergebnis wird in *statusMap* gespeichert.

2.2.2.5 _getAndDisplay()

Die Ergebnisse werden auf dem Bildschirm angezeigt.

2.2.2.6 _export()

Die ausgewählten Attribute werden in eine JSON-Datei exportiert. Dafür wird die Klasse *api_service* benötigt.

2.2.3 api_service

Die Klasse hat die eine Methode *exportJson(String domain, Map<String, dynamic> data)* und soll für jedes Attribut eine JSON-Datei erstellen und speichern. *String domain* enthält einen gültigen Domain-Namen einer Firma und *data* enthält die Ergebnisse der analysierten Attribute. Für jedes Element der Map *data* soll die vordefinierte Methode *description* aufgerufen und die darin enthaltene Beschreibung wird dem entnommen. Anschliessen wird eine neue Map erzeugt und in einen JSON-String umgewandelt. Das JSON-String wird in Bytes umgewandelt und gespeichert.

2.2.4 plugin_manager

Wenn im *home_screen* die Methode *_scan()* aufgerufen wird, dann wird in der Klasse *plugin_manager* kontrolliert, welches Python-Skript für die jeweilige Aufgabe ausgeführt werden soll und dieses wird dementsprechend aufgerufen. Genauer wird das im Kapitel 2.3 Plugin-System beschrieben.

2.2.5 cerificate_plugin

Beispielsweise sollen alle Zertifikate der Domain herausgefunden werden. Dann wird *plugin_manager* das Skript *certificate_plugin* ausführen. Wie für jedes Plugin ist *BasePlugin* die abstrakte Klasse von *certificate_plugin*. Sie gibt vor, dass die abstrakten Methoden *setup()*, *scan()* und *get()* auf ihre Art und Weise implementiert werden müssen. *Scan()* wird ausgeführt, um neue Daten zu suchen und *get()* dafür, um schon bestehende Daten in der Datenbank herauszuholen. Genauer wird das im Kapitel 2.3 Plugin-System beschrieben.

2.3 Plugin-System

Zur Realisierung eines möglichst modularen Systems, dessen Erweiterbarkeit sichergestellt werden soll, wurde ein Plugin-System entwickelt.

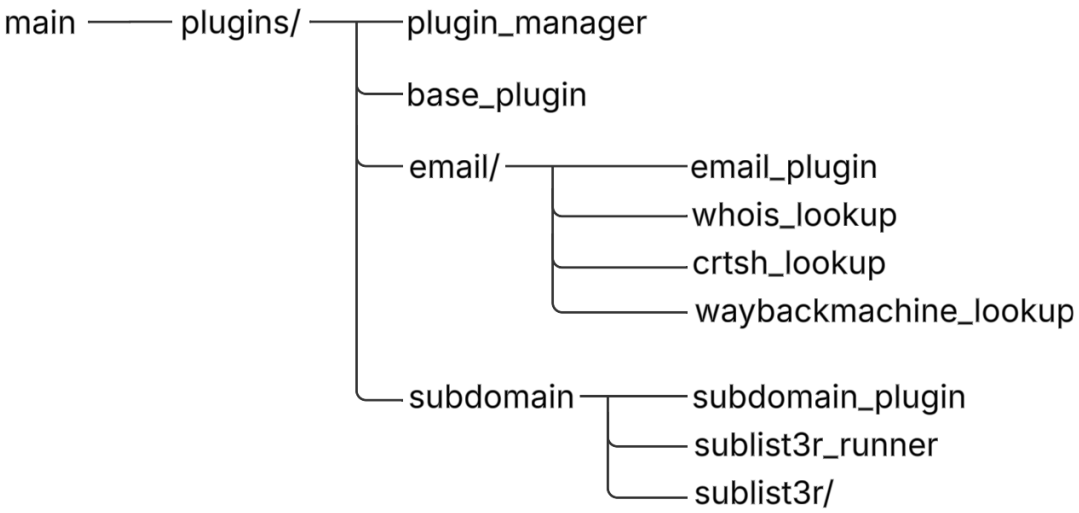

ABBILDUNG 2.3: Pluginsystem

Im Hauptprogramm *main* wird die Klasse des Pluginmanagers aufgerufen. Die Initialisierung aller Plugins erfolgt durch den Pluginmanager. Die Implementierung einer Superklasse, das BasePlugin, gewährleistet die Konsistenz aller Plugins.
Die Setup-Methode jedes Plugins überprüft, ob die Tabelle der Attribute innerhalb der Datenbank initialisiert wurde. Sollte eine Initialisierung noch nicht erfolgt sein, wird diese nachgeholt.
Des Weiteren sollten die Methoden *scan* und *get* implementiert werden, welche von der API-Call des Frontends aufgerufen werden. Die dritte API-Call, *describe*, ist bei jedem Plugin gleich und wird nur von BasePlugin bestimmt.

Jedes Attribut ist in einem spezifischen Ordner innerhalb des Plugin-Ordners abgelegt. In dem Ordner sind jeweils die Klasse *Plugin* des Attributs sowie mindestens eine weitere Klasse zu finden. Lookup-Klassen durchsuchen jeweils eine öffentliche Datenbank direkt, während Runner-Klassen ein externes Pytonscript von Github verwenden, um auf andere öffentliche Datenbanken zuzugreifen. Dies bedeutet, dass es mehrere Lookup-Klassen geben kann, die auf die gleiche öffentliche Datenbank zugreifen, jedoch unterschiedliche Daten daraus extrahieren. Die sehr hohe Modularität und Erweiterbarkeit des Systems geht mit dem Nachteil einher, dass die gleiche öffentliche Datenbank mehrfach aufgerufen wird.
Um die Struktur der Webseite noch weiter zu optimieren, wurde eine separate Tabelle für jedes Attribut angelegt. Somit wird bei jedem neu erstellten Attribut-Plugin automatisch eine neue Tabelle erstellt, ohne dass die bestehenden Informationen der anderen Attribute in der Datenbank beeinträchtigt werden.

Kapitel 3

Implementation

Zunächst werden die Werkzeuge vorgestellt, die zur Entwicklung der Applikation verwendet wurden und deren Funktionalität sicherstellen. Im zweiten Abschnitt folgt eine detaillierte Betrachtung der umgesetzten Attribute, wobei der Fokus auf den jeweils verwendeten Tools liegt, die für die Datenerhebung pro Attribut essenziell sind. Abschliessend werden Screenshots der Web-Oberfläche präsentiert und kurz erläutert, um einen Einblick in die visuelle Darstellung zu geben.

3.1 Werkzeuge

In diesem Kapitel werden die Werkzeuge für die Umsetzung eines Digital Twins näher erläutert. Flutter ist ein Framework, mit der die Webseite aufgesetzt wurde. Mit Python werden im Backend die Attribute ermittelt. Docker ist eine Container-Technologie, welche die Datenbank (PostgreSQL) und die ganzen Python Skripts beinhaltet.

3.1.1 Flutter

Flutter ist ein open-source-Framework, welches von Google entwickelt wurde. Es verwendet die Programmiersprache Dart.
Das Novum, welches Flutter anbietet, ist die Entwicklung von Cross-Platform-Apps. Das bedeutet, dass eine lauffähige App mit einem Code auf den Umgebungen Web, Android, iOS, Windows, Linux und macOS funktionieren kann. Zudem bietet Flutter eine reichhaltige Anzahl an Widgets, die das Erstellen von benutzerdefinierten Benutzeroberflächen sehr erleichtert. [12]
Obwohl Flutter sich durch Flexibilität und eine benutzerfreundliche Bedienung auszeichnet, gilt es zu bedenken, dass es sich um eine relativ junge Programmiersprache handelt, die erst seit 2017 existiert. Aus den genannten Gründen ist die Menge an Informationen im Internet zum Thema noch vergleichsweise gering, insbesondere im Vergleich zu etablierten Sprachen wie Java oder Javascript.

3.1.2 Docker

Docker ist eine Container-Technologie. Es ist eine freie Software und dient der Isolierung von Applikationen. Die Container beinhalten alle benötigten Pakete für das Funktionieren einer Applikation. [13]

Die wichtigsten Begriffe von Docker:

- Image: ausführbare Applikation als Artefakt, beinhaltet Scripts (Code), Dockerfile (Konfiguration). Hier befindet sich das lauffähige Programm.
- Container: Ein Image, das ausgeführt wird, ist ein Container. Es können mehrere Container aus einem Image erstellt werden. Wenn das Programm fertig ist, dann wird der Container beendet.
- Dockerfile: Konfigurationsdaten für ein lauffähiges Image. Hier werden die einzelnen Zeilen ausgeführt.
- Registry: Eine Verwaltung von 3th Party Images.

3.1.2.1 Unterschied eines Dockers zu einer virtuellen Maschine

Es gibt drei verschiedene Layer, in die man einen funktionieren Computer unterteilen kann. Hardware, OS Kernel (Operating System = Betriebssystem) und OS Application Layer. Docker Container bildet nur den OS Application Layer ab und die VM den OS Kernel und die OS Application Layer. VM simuliert somit ein ganzes Betriebssystem. Diese Unterschiede ergeben viele Vorteile für den Docker Container, die im Kapitel 3.1.2.2 näher beleuchtet werden. [14]

3.1.2.2 Probleme die Docker löst und seine Vorteile

Docker Container weisen mehrere Vorteile gegenüber VMs (Virtual Machine) auf. Sie verbrauchen aufgrund ihrer technischen Unterschiede viel weniger Ressourcen. Sie sind flexibler einsetzbar und effektiver. VMs beinhalten alle erforderlichen Binärdateien und Bibliotheken, die mehrere Gigabytes gross sind. VMs können ausserdem nur langsam booten. Docker Container dagegen haben nur eine Grösse von mehreren Megabytes.

Dafür sind VMs mit allen Betriebssystemen kompatibel. Docker Container dagegen ist nur Linux-kompatibel.

Ein Computer mit einem Windows-Kernel und einem Windows-Application-Layer kann keine Linux-basierten Docker Container direkt ausführen, da der Docker Container immer auf einen Linux Kernel basiert. Um dieses Problem zu lösen, wurde Docker Desktop für Windows und macOS entwickelt. Es ermöglicht, Docker Container auch auf Systemen mit nicht-Linux-Kernel auszuführen, indem es eine virtuelle Linux-Umgebung bereitstellt.

Ein weiterer grosser Vorteil von Docker ist die Plattformunabhängigkeit: Anwendungen können in beliebigen Umgebungen ausgeführt werden, ohne dass diese Umgebungen speziell konfiguriert oder angepasst werden müssen.

3.1.2.3 Wichtige Docker Commands

- docker images -> Zeigt alle Images.
- docker ps -> Zeigt alle laufenden Container.
- docker ps -a -> Zeigt alle Container.
- docker rm [„name“] -> Löscht den spezifischen Container.
- docker image rm [„name“] -> Löscht das spezifische Image.
- docker build -t [python_ip_adress] . -> Baut ein Image von Dockerfile und Scripts.
- docker run –name [ip_adress] [python_ip_adress] -> Baut einen Container [ip_adress] von image.
- docker restart ip_adress -> Startet den beendeten Container neu.
- docker logs ip_adress -> Zeigt alle Logs des Containers.
- docker-compose up –build -> Baut alle Images neu und startet alle Container basierend auf dem docker-compose.yml.

3.1.3 Python Scripts

Python Script bilden die Grundlage für das Ermitteln der Ergebnisse der Attribute. Alle Scripts befinden sich im Ordner Backend und wurden mit der Programmiersprache Python verfasst. Sie sind für das Ermitteln von IP-Adressen, E-Mails, Subdomains, usw. verantwortlich. Jedes Attribut benötigt zusätzlich ein Plugin-Skript. Mit diesen ist es möglich, eine Datenbank aufzusetzen, Ergebnisse zu ermitteln und Ergebnisse in der Datenbank zu speichern. Im Kapitel 3.2 werden die Attribute näher beleuchtet.

3.2 Umsetzung

Das vorliegende Kapitel „Umsetzung“ gibt Aufschluss über die Implementierung der Attribute, den Aufbau des Webdesigns sowie die Art und Weise der Datenspeicherung in der Datenbank.

3.2.1 Attribute

Die Grundbausteine eines Digital Twins werden durch seine Attribute definiert. Ein Attribut kann beispielsweise die Subdomains einer Hauptdomain umfassen. Die Identifizierung von Attributen einer Firma aus öffentlichen Informationen erfordert verschiedene Methoden, die in einigen Fällen auch kombiniert werden müssen.
Die Ermittlung relevanter Attribute erfolgt unter anderem mithilfe von Hack The Box Academy, insbesondere des Penetration Tester Role Path. [15] Auch Google und ChatGPT wurden als Informationsquellen genutzt.
Die vorliegende Tabelle präsentiert eine vollständige Übersicht über die identifizierten Attribute sowie die zugrunde liegende Technik.

Attribut	Tool
DNS-Records	Dig Befehl
Subdomains	sublist3r
Zertifikate	Crt.sh
Endpunkte	Wayback Machine
(Geleakte) E-Mails	Wayback Machine, Crt.sh, Whois, Dehashed
(Geleakte) Telefonnummern	Wayback Machine, Whois und Dehashed
Dienste	Shodan

In den nachfolgenden Unterkapiteln erfolgt eine detaillierte Erläuterung der Attribute. Dabei werden die nachstehenden Fragen zu jedem Attribut erörtert.

- Was ist das Attribut?
- Warum ist dieses Attribut aus cyber-security-technischer Sicht wichtig?
- Aus welchen öffentlichen Datenbanken wird dieses Attribut extrahiert?

3.2.1.1 DNS-Records

Ein DNS-Record ist ein Eintrag in einer DNS-Zonen-Datei, die bestimmte Informationen zu einer Domain enthält. DNS steht für „Domain Name System”. Jeder Record-Typ erfüllt eine bestimmte Funktion, beispielsweise die Zuordnung von Mailservern oder die Adressauflösung.
Aus sicherheitstechnischer Sicht sind DNS-Records wichtig, da sie Informationen zu Subdomains, Mail- und Nameservern anzeigen können. Auch kann man die Veränderungen des Systems über die Zeit hinweg nachvollziehen. So kann beispielsweise erkannt werden, wenn eine neue Subdomain wie test.payment.example.ch live geht, die einen neuen Angriffspunkt anzeigt. Auch die Infrastruktur eines Netzwerkes kann aufgezeigt werden. Beispielsweise können Nameserver den Hosting-Provider verraten. Somit können mögliche Schwachstellen sowie der Datenfluss erfasst werden.
In den weiteren Abschnitten wird der Nutzen der verschiedenen DNS-Records genau erläutert. [16]

Alle folgenden DNS-Records wurden über das Dig-Kommandozeilen-Tool auf der Webseite des Digital Twin gefunden. Dieser Befehl wird genutzt, um verschiedene DNS-Record-Abfragen direkt an den DNS-Server zu senden. Sie sind nützlich für die Netzwerk-Analyse.
Abgesehen von PTR lautet der Befehl für alle genutzten Records gleich. AAAA ist beispielsweise die Abkürzung für den AAAA-Record.

- dig +short example.ch [DNS-Record-Abkürzung]

Für den PTR-Record wird der folgende Befehl mit der dazugehörenden IP-Adresse verwendet:

- dig +short -x [IP-Adresse]

[17]

3.2.1.1.1 A-Records

Der vollständige Name des A-Records lautet „Address Record". Er zeigt vom Hostname auf die IPv4-Adresse. [16]
Daher wird jeder A-Record-Eintrag nur mit zwei Eigenschaften gespeichert. Die erste ist die IPv4-Adresse selbst, die zweite das Scan-Datum. Dieses Datum zeigt an, wann dieser Eintrag im Digital Twin eingelesen wurde. Anhand dieses Datums kann entschieden werden, ob sich ein neuer Scan lohnt oder ob die vorhandenen Daten ausreichen. Nachfolgend ist ein Beispiel eines Eintrags von zhaw.ch zu sehen.

IPv4-Adresse	In Digitalem Twin gescannt am
160.85.192.183	2025-05-23

TABELLE 3.1: Beispielhafter Eintrag für einen A-Record

Wenn eine angreifende Person die IPv4-Adresse einer Webseite kennt, kann sie Nmap nutzen. Dies ist ein leistungsstarkes Tool für die Netzwerkanalyse. Damit kann man nach offenen, gefilterten sowie geschlossenen Ports suchen. Auch kann man gleichzeitig die laufenden Dienste und Versionen bestimmen. [18]
Mit diesen Informationen kann die angreifende Person Metasploit nutzen, um nach Schwachstellen zu suchen. Metasploit ist ein Framework, das verschiedene Module bietet, um Exploits, Payloads und Scans zu nutzen. Dabei können bekannte Sicherheitslücken automatisiert ausgenutzt oder eigene Module für einen Angriff entwickelt werden. [19]

3.2.1.1.2 AAAA-Records

Um einen Hostnamen auf eine IPv6-Adresse anzuzeigen, ist das AAAA-Record erforderlich, das im vollen Namen IPv6-Record heisst. [16]
Da die ZHAW keine IPv6-Adresse hat, geben wir hier das Beispiel der 3-plan.ch an.

IPv6-Adresse	In Digitalem Twin gescannt am
2a01:ab20:0:4::88	2025-05-14

TABELLE 3.2: Beispielhafter Eintrag für einen AAAA-Record

Die sicherheitsrelevanten Aspekte für die IPv6-Adressen sind identisch mit den Aspekten des Abschnitts A-Record für IPv4-Adressen und werden daher an dieser Stelle nicht erneut genauer erläutert.

3.2.1.1.3 MX-Records

MX-Record steht für *Mail Exchange Record* und spezifiziert den Mail-Server, welcher für die Handhabung der E-Mails für die Domain verantwortlich ist. [16]
Hier ist nochmal ein Beispiel der Domain zhaw.ch:
Ein MX-Record setzt sich aus dem Mailservernamen sowie der Präferenz zusammen. Die Ausprägung einer niedrigen Präferenz signalisiert demnach eine höhere Priorität und resultiert in einer vorherigen Kontaktaufnahme. Der Begriff „Mailservername" bezeichnet den Hostnamen des Servers. In dem vorliegenden Beispiel wird durch den Mailservernamen der Name des von Microsoft entwickelten Programms „Outlook" verraten. [20]

Präferenz	Mailserver (MX)	In Digitalem Twin gescannt am
10	zhaw-ch.mail.protection.outlook.com.	2025-05-23

TABELLE 3.3: Beispielhafter Eintrag für einen MX-Record

Die MX-Records sind von Relevanz, da die Person, die den Angriff ausführt, in ihnen erkennen kann, ob die MX-Einträge fehlerhaft, veraltet oder lückenhaft sind. Es besteht demnach die Möglichkeit, Phishing- und Spoofing-Angriffe vorzubereiten, den Mailverkehr umzuleiten oder Backup-Server auszuspähen. [21]

3.2.1.1.4 NS-Records

Das NS-Record, eine Abkürzung für *Name Server Record*, fungiert als Vermittler zwischen einer Domain und einem spezifischen authoritativen Nameserver innerhalb einer bestimmten DNS-Zone. [16]
Im Folgenden wird ein weiteres Beispiel von zhaw.ch präsentiert.

Name Server	In Digitalem Twin gescannt am
dns1.zhaw.ch.	2025-05-23

TABELLE 3.4: Beispielhafter Eintrag für einen NS-Record

Der Name Server erweist sich als ein nützliches Instrument, um einen Zone Transfer auszulösen. Für den Transfer ist der Name des Nameservers erforderlich. In der Regel wird er genutzt, um eine vollständige Kopie der DNS-Zone von einem primären zu einem sekundären Nameserver zu übertragen. Zur Erhöhung der Redundanz und der Fehlertoleranz erfolgt die regelmässige automatische Übertragung sämtlicher DNS-Records an autorisierte sekundäre Server.
Wird der Nameserver jedoch nicht adäquat konfiguriert, besteht für die angreifende Person die Möglichkeit, einen Zonentransfer zu initiieren. Die vorliegende Persona ist folglich dazu befähigt, auch Informationen zu erlangen, die nicht öffentlich bekannt sind. [22]

3.2.1.1.5 PTR-Records

Ein PTR-Record fungiert als Instrument zur Durchführung eines Reverse DNS Lookups, dessen Funktion in der Umwandlung einer IP-Adresse in einen Hostnamen besteht. PTR-Record ist dabei die Abkürzung für *Pointer Record*. [16]
Nachstehend wird ein weiteres Beispiel von zhaw.ch präsentiert.

IP-Adresse	PTR-Domain	In Digitalem Twin gescannt am
160.85.192.80	srv-dc-102.zhaw.ch	2025-05-23

TABELLE 3.5: Beispielhafter Eintrag für einen PTR-Record

Nach Ermittlung einer IP-Adresse besteht die Möglichkeit, mittels eines PTR-Record Lookups weitere Server und deren Funktion zu ermitteln. Es besteht die Möglichkeit, dass ein VPN-Zugang oder eine Entwickler-API endeckt werden, welche sonst schwer aufzufinden sind. [23] Dies ist für die angreifende Person von signifikanter Relevanz, um eine Analyse der Netzwerkstruktur zu ermöglichen. Die so gewonnenen Erkenntnisse können im weiteren Verlauf bei einem Exploit oder einem Social Engineering Angriff genutzt werden.

3.2.1.1.6 SOA-Records

Ein SOA-Record dient der Spezifikation administrativer Informationen bezüglich der DNS-Zone. Dazu gehören unter anderem der Name des primären Nameservers sowie die E-Mail-Adresse des Administrators. Die Abkürzung „SOA-Record“ steht für „Start of Authority Record“. [16]
Nachfolgend wird ein weiteres Beispiel von zhaw.ch präsentiert. Die Zonen-Daten umfassen sämtliche DNS-Records einer Zone sowie deren Verwaltungsinformationen.
Die Eigenschaften des SOA Records präsentieren sich wie folgt: Der Parameter *Expire* spezifiziert die Zeitspanne, innerhalb derer ein Sekundärserver DNS-Antworten auf Basis veralteter Zonen-Daten bereitstellen darf, sofern der Primärserver nicht erreichbar ist. Nach Ablauf dieser Zeit werden die Zonendaten als veraltet betrachtet, und der Sekundärserver stellt den Dienst ein, falls kein erfolgreicher Zonentransfer erfolgt ist. Die Versionsnummer der Zonedaten wird als „Serial“ bezeichnet. Der Parameter *Refresh* spezifiziert das Zeitintervall (in Sekunden), innerhalb dessen ein Sekundärserver die Änderung der Serialnummer beim Primärserver prüft. Im Falle einer Änderung dieser Versionsnummer wird dem Sekundärserver signalisiert, dass ein Update per Zonentransfer erforderlich ist. Der Begriff „Retry“ bezeichnet das Zeitintervall, nach dem ein Sekundärserver nach einem fehlgeschlagenen Kontaktversuch erneut den Versuch unternimmt, den Primärserver zu erreichen. Es handelt sich dabei um eine Fehlerbehandlungsstrategie. [24]

Retry	Expire	Serial	Refresh
300	604800	2731581971	1800

TABELLE 3.6: Beispielhafter erster Teil eines Eintrages für einen SOA-Record

Die E-Mail-Adresse des Administrators lautet hostmaster@zhaw.ch, wobei das Punkt-Zeichen als Zeichen für das @ zu verstehen ist.
Die „Minimum TTL“ im SOA-Record definiert die Zeitspanne, innerhalb derer ein DNS-Resolver negative Antworten (beispielsweise bei nicht vorhandenen Domainnamen) zwischenspeichern darf, bevor erneut eine Anfrage zu stellen ist.

Hostmaster	Minimum TTL
hostmaster.zhaw.ch.	68400

TABELLE 3.7: Beispielhafter zweiter Teil eines Eintrages für einen SOA-Record

Der Primary Nameserver fungiert als primäre Informationsquelle für die sekundären Server in Bezug auf die Zoneninformationen. [24]

Primary Nameserver	In Digitalem Twin gescannt am
dns1.zhaw.ch	2025-05-23

TABELLE 3.8: Beispielhafter zweiter Teil eines Eintrages für einen SOA-Record

Die Informationen sind für die angreifende Person nützlich, um Fehlkonfigurationen zu identifizieren und um den Namen des Nameservers in Erfahrung zu bringen. Dieser ist wiederum relevant für einen *Zone Transfer*. Des Weiteren besteht die Möglichkeit, dass die E-Mail-Adresse des Administrators für Spearphishing-Angriffe instrumentalisiert wird. Aufgrund der Tatsache, dass die Wahrscheinlichkeit besteht, dass die E-Mail-Adresse auch für andere Logins genutzt wurde, besteht die Möglichkeit, dass diese auch für das sogenannte *Passwordbruteforcing* missbraucht wird.

3.2.1.1.7 TXT-Records

Das TXT-Record steht für *Text Record* und speichert verschiedene Textinformationen, welche oft für Domainverifizierungen genutzt werden.[16]
Nachfolgend wird ein Beispiel der Webseite 3-plan.ch präsentiert. Der Text Record `"MS=ms49296141"` könnte ein Indikator für ein Microsoft Office-Produkt und ein Token sein. Dieser Aspekt stellt einen signifikanten Faktor in Bezug auf die Informationsbeschaffung für eine angreifende Person dar.

Text Record	In Digitalem Twin gescannt am
`"MS=ms49296141"`	2025-05-23

TABELLE 3.9: Beispielhafter Eintrag für einen TXT-Record

Ein TXT-Record mit dem Inhalt `"1password=..."` ist ein gutes Beispiel dafür, weshalb TXT-Records sicherheitsrelevant sind. Dieser TXT-Record zeigt höchstwahrscheinlich ein Passwort an. Zudem besteht die Möglichkeit, dass die Organisation dasselbe Passwort über verschiedene Infrastrukturen hinweg verwendet. Dies ist für einen Angreifer oder eine Angreiferin sehr interessant. Angenommen das Passwort

steht nicht im Klartext sondern als ein Token, so kann man ihn nicht direkt einsetzen. Aber es ist dennoch möglich, es für Social Engineering Attacken oder gezielte Phishing-Kampagnen zu nutzen.[16]

3.2.1.2 Subdomains

Eine Subdomain stellt einen Teil einer URL dar. Der Einsatz von Subdomains zur Ergänzung einer Hauptdomain kann verschiedene Ziele verfolgen und vielfältige Vorteile mit sich bringen. Es besteht die Möglichkeit, verschiedene Segmente einer Webseite zu separieren. Ein Beispiel für derartige Plattformen sind virtuelle Verkaufsräume, Diskussionsforen oder Weblogs. Des Weiteren besteht die Möglichkeit, eine lokale oder sprachspezifische URL zu erstellen oder eine spezielle Subdomain ausschliesslich für mobile Endgeräte zu generieren.[25]
Im Folgenden werden verschiedene Beispiele aufgeführt.

- blog.yoursite.com
- forum.yoursite.com
- shop.yoursite.com
- shop.de.yoursite.com
- shop.ch.yoursite.com
- mobile.blog.yoursite.com

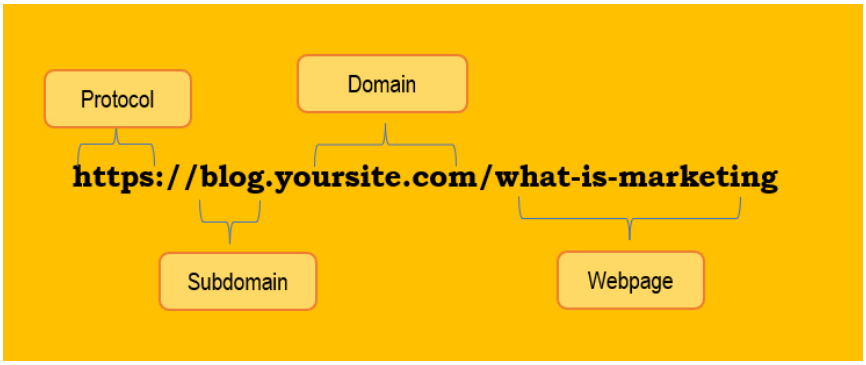

ABBILDUNG 3.1: URL-Struktur [25]

Die Subdomain wird dabei vor der Domain platziert. [25]
In der vorliegenden Arbeit wird vereinfacht der Begriff „Subdomain“ für die Kombination aus Subdomain und Domain verwendet.
Im Folgenden wird ein weiteres Beispiel der Webseite zhaw.ch präsentiert, welches die Existenz einer Domain veranschaulicht, die mutmasslich zum Testen eines Kreditkartensystems aus der Perspektive eines Administrators genutzt wurde. Wird eine Domain hauptsächlich von einer Administratorin oder einem Administrator genutzt, besteht die Möglichkeit je nach Konfiguration, dass nach dem Eindringen in die Domain automatisch Administratorinnen- oder Administratorenrechte erworben werden könnten.

Subdomainname	In Digitalem Twin gescannt am
cardservice-test-admin.zhaw.ch	2025-05-23

TABELLE 3.10: Beispielhafter Eintrag für eine Subdomain

Für den Digital Twin sind die Domains von signifikanter Relevanz, da sie potenzielle Angriffsflächen bieten. Unternehmen setzen Subdomains, wie beispielsweise „test.example.com“ oder „development.example.com“ ein, um neue Funktionalitäten zu entwickeln und zu testen. Des Weiteren besteht die Möglichkeit, auf nicht öffentlich einsehbare, versteckte Login-Portale zuzugreifen. Es besteht zudem die Möglichkeit, veraltete Webanwendungen zu identifizieren, die mit unzeitgemässer Software ausgestattet sind und bekannte Sicherheitslücken aufweisen. [26]

Für die Suche der Subdomains wird in dieser Bachelorarbeit das Programm Sublist3r verwendet. Die entsprechende Publikation ist auf der Plattform Github verfügbar und wurde durch den Nutzer aboul3la bereitgestellt. Die folgenden öffentlichen Quellen werden von Sublist3r genutzt:

- Suchmaschinen: Google, Yahoo, Bing, Baidu, Ask
- Dienste: Netcraft, Virustotal, ThreatCrowd, DNSdumpste, PassiveDNS, crt.sh

[27]

3.2.1.3 Zertifikate

Die vorliegende Tabelle veranschaulicht ein weiteres Beispiel der zhaw.ch. Die Spalten „Gültig ab“ und „Bis“ indizieren die Gültigkeitsdauer eines Zertifikats.
Die Angabe „Ausgestellt von“ verweist auf die Zertifizierungsstelle, von der das jeweilige Zertifikat ausgestellt wurde. Es ist essenziell, die Domain bzw. Subdomain zu spezifizieren, da das Zertifikat nur für diese Domain Gültigkeit besitzt.

Gültig ab	Gültig bis	Ausgestellt von	Domain/Subdomain
2021-03-31	2021-06-29	C=US, O=Let’s Encrypt, CN=R3	my-test1.zhaw.ch

TABELLE 3.11: Beispielhafter erster Teil eines Eintrages für ein Zertifikat

„In öffentlicher Datenbank erfasst am“ zeigt den Zeitpunkt an, zu dem eine Information in der öffentlichen Datenbank gespeichert wurde, aus der die Webseite Digitaler Twin ihre Daten bezieht. Dies ermöglicht eine bessere Identifikation von veralteten, aber immer noch wichtigen, Informationen. Die Wildcard-Abfrage zeigt an, ob das Zertifikat für alle Subdomains auf einer Ebene zuständig ist. Beispielsweise ist *.example.ch für www.example.ch sowie mail.example.ch gültig aber nicht für example.ch sowie forum.test.example.ch.

Wildcard-Zertifikat (Ja/Nein)	In öffentlicher Datenbank erfasst am
Nein	2021-04-01

TABELLE 3.12: Beispielhafter zweiter Teil eines Eintrages für ein Zertifikat

In Digitalem Twin gescannt am
2025-05-23

TABELLE 3.13: Beispielhafter dritter Teil eines Eintrages für ein Zertifikat

Die Daten der Zertifikate wurden auf der Seite Crt.sh aufgefunden. Die vorliegende Datenbank speichert in regelmässigen Abständen Zertifikate, sodass auch abgelaufene Zertifikate eingesehen werden können.[28]
Die Webseite Digitale Twin lädt beim Scan über eine kostenlose API die Informationen von Crt.sh herunter und speichert diese in der Datenbank. Für Angreifer und Angreiferinnen lohnt es sich aus verschiedenen Gründen, die Zertifikate zu überprüfen. Ein Grund sind die Sicherheitsrisiken, die mit Wildcard-Zertifikaten einhergehen. Die Problematik, die sich bei der gemeinsamen Nutzung eines Wildcard-Zertifikats stellt, ist die, dass im Falle einer Kompromittierung eines Servers das Zertifikat entwendet werden kann. Eine angreifende Person könnte dieses Wissen dazu nutzen, um Subdomains, wie beispielsweise admin.example.ch, täuschend echt nachzubilden und damit Personen mit dieser Phishing-Webseite blenden. [29]

3.2.1.4 Endpunkte

In dieser Arbeit bezeichnen wir die URL als Endpunkte. Am Anfang der URL steht immer das Protokoll HTTP in seiner unsicheren Variante und HTTPS in seiner sicheren Variante. Danach kommt jeweils die Domain sowie eine etwaige Subdomain. Zum Schluss erreicht der Pfad im Server zur gewünschten Ressource. Diese Ressourcen könnten HTML-Seiten aber auch PDsF sein. [30]

Ein weiteres Beispiel stammt von der albl.ch. Hier kann man die Administratorenseite für die Webseite finden. Für einen Angreifer oder eine Angreiferin ist das ein möglicher Angriffspunkt.

URL
https://albl.ch/wp-admin

TABELLE 3.14: Beispielhafte erste Hälfte eines Eintrages für einen Endpunkt

In öffentlicher Datenbank erfasst am	In Digitalem Twin gescannt am
2022-09-21	2025-05-14

TABELLE 3.15: Beispielhafte zweite Hälfte eines Eintrages für einen Endpunkt

Aus der Perspektive eines potenziellen Angreifers kann das systematische Sammeln von URLs aus verschiedenen Gründen lohnenswert sein. Eine besondere Problematik besteht in der Nutzung von URLs, die das unverschlüsselte HTTP-Protokoll einsetzen. Es existieren verschiedene Möglichkeiten, diese zu realisieren. Es folgt eine Aufzählung von Punkten.

- Distributed Denial of Service (DDoS): Überlastung des Servers durch massenhafte parallele Anfragen.
- Cross-Site Scripting (XSS): Einschleusen eines JavaScript-Codes zur Übernahme von Benutzersitzungen oder zum Diebstahl von Cookies.
- SQL-Injection: Einschleusen eines schädlichen SQL-Codes zur Manipulation von Datenbanken.
- Cross-Site Request Forgery (CSRF): Ausführen unerwünschter Aktionen im Kontext eines eingeloggten Nutzers.

Die Tatsache, dass HTTP keine gesicherte Kommunikation ermöglichen, begünstigt diese Art von Angriffen, da die Kommunikation im Klartext über das TCP-Protokoll erfolgt. Es besteht demnach die Möglichkeit, dass auch auf Netzwerkebene gezielte Angriffe stattfinden. HTTP-Flood-Angriffe stellen eine besondere Herausforderung dar. Es handelt sich um legitime HTTP-Anfragen, die in einer solchen Menge und mit einer derartigen Präzision eingesetzt werden, dass sie schwer zu erkennen und zu blockieren sind. [31]
Über diese URLs kann ein Angreifer oder eine Angreiferin auch die technologische Entwicklung und potenzielle Schwachstellen rückverfolgen. Dabei können auch eventuell noch vergessene URLs oder Dateien enthalten sein, die sensible Informationen beinhalten. Diese Informationen können auch für einen potenziellen Exploit oder eine Social Engineering Attacke genutzt werden. [32]
Die Webseite Digitaler Twin zieht diese Daten aus der öffentlichen, kostenlosen API der Wayback Machine.[33] Die „Wayback Machine" ist eine Webseite, die sich seit 1996 automatisch durch das Internet „crawlt". Sie speichert vollständige Schnappschüsse von Webseiten inklusive ihrer Bilder, CSS, HTML, usw. Mithilfe dieser Schnappschüsse ist der Zugriff auf ältere Versionen einer Webseite möglich, auch wenn sich die Webseite in der Zwischenzeit geändert hat. Beliebtere Webseiten, also öfter besuchte Webseiten, werden häufiger gescannt und haben daher mehr aktuelle Snapshots. Dieses Prinzip gilt auch für die einzelnen HTML-Seiten einer Webseite.[32]

3.2.1.5 E-Mails

Hier ist ein Beispiel für gefundene E-Mails der Domain zhaw.ch. Es wird einerseits die gefundene E-Mail angezeigt, als auch überprüft, ob ein Klartext Password mit in Zusammenhang der E-Mail im Internet veröffentlicht wurde.

E-Mail	Password geleaked (Ja/Nein)
peter.schwendner@zhaw.ch	Nein

TABELLE 3.16: Beispielhafte erste Hälfte eines Eintrages für eine E-Mail

Wenn die öffentliche Datenbank selbst kein Datum speichert, wird hier immer das aktuelle Datum des Scans genommen.

In öffentlicher Datenbank erfasst am	In Digitalem Twin gescannt am
2025-05-19	2025-05-23

TABELLE 3.17: Beispielhafte zweite Hälfte eines Eintrages für eine E-Mail

Der Digitale Twin scannt E-Mails aus drei verschiedenen Quellen. Die erste Quelle ist Crt.sh, da E-Mails auch in Zertifikaten enthalten sein können. [28]
Auch werden die URLs der Wayback Machine genutzt, um über die Wayback Machine alle gültigen URLs abzusuchen und mithilfe eines Regex nach E-Mails zu suchen.[33] Dabei ist es wichtig zu wissen, dass wir nur gültige E-Mail-Endungen beachten und dass example@example.pdf nicht als E-Mail gezählt wird, da offensichtlich .pdf keine gültige Endung ist. Diese Liste der Internet Assigned Numbers Authority (IANA) zeigt alle Top-Level-Domains an. Die gefundenen E-Mails müssen also mit einer dieser Endungen enden, um als gültig zu gelten. [34]
Als letzte Quelle wird Whois genutzt. Whois ist sozusagen das digitale Telefonbuch des Internets. Über Whois können wichtige Daten wie Domainname, Adminkontakte, Erstellungsdatum und vieles mehr abgefragt werden. Für diese Arbeit sind die Adminkontakte relevant, die eine E-Mail-Adresse beinhalten können. Diese E-Mail-Adressen sind für die Identifizierung von Schlüsselpersonen relevant und ermöglichen Phishing und Social Engineering Angriffe. [35]
Bei diesem Attribut wird auch überprüft, ob ein Klartextpasswort im Zusammenhang mit der E-Mail geleakt wurde. Diese Informationen erhalten wir von der Dehashed-API.[36] Diese sucht verschiedene Datenbanken im Darknet auf, sammelt die Daten und bereitet sie in einer API zum Abruf vor. Die Preisgestaltung wurde während der Ausführung der Bachelorarbeit von Dehashed überarbeitet und ist nun teurer. Um die API zu nutzen, muss man einerseits einen monatlichen Zugang zur API kaufen und andererseits Kredite erwerben, um einzelne API-Abfragen zu starten. Pro Abfrage wird ein Kredit fällig. Dabei ist es egal, ob eine einzelne E-Mail-Abfrage oder eine ganze Domainabfrage durchgeführt wird.

Für diese Bachelorarbeit wurde ein zweimonatiger API-Zugang bezahlt.

Kosten Pro Monat	Kosten für Bachelorarbeit
19.42 CHF	38.83 CHF

TABELLE 3.18: Preisgestaltung von Dehashed für die monatliche Nutzung der API

Für dieses Projekt wurden zudem 2'100 Kredits eingekauft.

Kosten Pro Kredit	Kosten für Bachelorarbeit
0.027 CHF	55.95 CHF

TABELLE 3.19: Preisgestaltung von Dehashed für die einzelnen Kredite

Insgesamt belaufen sich diese Kosten somit auf 94.78 CHF.
Laut einer Umfrage von Forbes aus dem Jahr 2024 verwenden 32% der Nutzer und Nutzerinnen dasselbe Passwort mehrfach. Wenn also ein Klartextpasswort gefunden wurde, kann man davon ausgehen, dass diese Person dieses Passwort mindestens einmal wiederverwendet hat. [37] Diese Passwort-Wiederverwendung kann ein Angreifer oder eine Angreiferin nutzen, um sich in verschiedene Accounts mit der gleichen E-Mail und gleichem Passwort einzuloggen.

3.2.1.6 Telefonnummern

Nachfolgend ist wieder ein beispielhafter Eintrag der zhaw.ch bezüglich den Telefonnummern abgebildet.

Telefonnummern	Password geleaked (Ja/Nein)
+41589344105	Nein

TABELLE 3.20: Beispielhafte erste Hälfte eines Eintrages für eine Telefonnummer

In öffentlicher Datenbank erfasst am	In Digitalem Twin gescannt am
2025-05-19	2025-05-19

TABELLE 3.21: Beispielhafte zweite Hälfte eines Eintrages für eine Telefonnummer

Der Digitale Twin scannt zwei verschiedene öffentliche Datenbanken. Wir scannen wieder über die Wayback Machine mithilfe eines Regex nach Telefonnummern. Wichtig dabei ist, dass momentan nur Schweizer Nummern berücksichtigt werden. [33]
Die letzte Quelle ist wie bei den E-Mails die Whois Datenbank. [35]
Die Securitygefahren sind dieselben wie für die E-Mails und werden daher nicht nochmal aufgelistet.
Gleich wie bei den E-Mails suchen wir mithilfe von Dehashed nach veröffentlichten Klartextpasswörtern. [36] Im Gegensatz zu den geleakten Passwörtern in Verbindung mit E-Mail-Adressen sind Passwörter in Verbindung mit Telefonnummern

anders nützlich. Sie werden weniger zum Einloggen verwendet, sondern eher für Social Engineering. Man könnte beispielsweise beim Helpdesk anrufen und mithilfe von Informationen wie alten Telefonnummern, Wohnadressen, Namen und ehemaligen Passwörtern versuchen, sich als die richtige Nutzende auszugeben.

3.2.1.7 Dienste

Dieses Attribut dient der Anzeige der auf dem jeweiligen Port ausgeführten Dienste. Des Weiteren wird die Version des Dienstes angezeigt, ebenso wie das verwendete Protokoll. Des Weiteren erfolgt die Anzeige der dazugehörigen IP-Adresse sowie der Domain.

Anbei finden Sie ein weiteres Beispiel von zhaw.ch.

Port	Dienst	Version	Protokoll	IPv4-Adresse	Domain/Subdomain
443	nginx	1.18.0	tcp	160.85.67.113	srv-lab-t-424.zhaw.ch

TABELLE 3.22: Beispielhafter erster Teil eines Eintrages für ein Zertifikat

Die Common Platform Enumeration (CPE) wird ebenfalls angezeigt. Bei dem vorliegenden Namensschema handelt es sich um eine standardisierte, maschinenlesbare Form der eindeutigen Identifikation von IT-Komponenten wie Hardware, Software und Betriebssystemen. [38]

Common Platform Enumeration (CPE)
`cpe:/a:f5:nginx:1.18.0, cpe:/o:canonical:ubuntu_linux, cpe:/o:linux:linux_kernel`

TABELLE 3.23: Beispielhafter zweiter Teil eines Eintrages für ein Zertifikat

Abschliessend wird die „Common Vulnerabilities and Exposures, kurz CVE“, angezeigt. Die CVE bezeichnet einen öffentlich dokumentierten Sicherheitsfehler in Software-Systemen. Jeder CVE-Eintrag ist mit einer eindeutigen CVE-ID verknüpft, die zur Erfassung der Schwachstelle dient. [39] Aus dem vorliegenden Beispiel geht hervor, dass der nginx-Dienst zum betreffenden Zeitpunkt drei bekannte Schwachstellen aufwies.

Common Vunerabilites and Exposures (CVE)
[CVE-CVE-2023-44487, CVE-CVE-2021-23017, CVE-CVE-2021-3618]

TABELLE 3.24: Beispielhafter dritter Teil eines Eintrages für ein Zertifikat

In Digitalem Twin gescannt am
2025-05-23

TABELLE 3.25: Beispielhafter vierter Teil eines Eintrages für ein Zertifikat

Mit diesen Informationen ist es möglich, Metasploit in gleicher Weise wie im Abschnitt über A-Records zu nutzen. Es kann explizit nach diesen CVE-Schwachstellen gesucht und diese anschliessend ausgenutzt werden.[19] Die Webseite Digitaler Twin hat die Informationen über die Dienste mithilfe der Shodan-API erhoben. [40]. Es besteht die Möglichkeit, dass die Informationen, über die Shodan verfügt, nicht mehr auf dem aktuellen Stand sind.
Shodan ist eine Suchmaschine, die sich auf Dienste spezialisiert hat und in ihrer Funktionsweise Google ähnelt. Zu den möglichen Diensten zählen Server, Router oder auch Kameras. Google durchsucht demnach Webseiten, während Shodan das gesamte Internet durchsucht. Das Tool wird unter anderem dafür genutzt, die Netzwerksicherheit zu verbessern oder eine Marktanalyse durchzuführen. In der vorliegenden Arbeit wird Shodan genutzt, um KMUs die Möglichkeit zu bieten, ihre Cyber-Risiken mithilfe der Webseite Digital Twin besser zu bewerten. Dies geschieht, indem die öffentlich zugänglichen Daten analysiert werden. Shodan bevorzugt beliebte, sicherheitsrelevante und stark genutzte Dienste. Daher kann es vorkommen, dass kleinere Firmen weniger stark gescannt werden als grosse. [41] Um die Shodan-API zu nutzen, ist ein monatlicher Zugang erforderlich. Für diese Arbeit konnte ein bereits vorhandener API-Zugang der Forschungsgruppe Information Security der ZHAW genutzt werden. Daher fielen im Rahmen dieser Arbeit keine zusätzlichen Kosten an. Der von der Forschungsgruppe genutzte monatliche Zugang kostet 69 Dollar. Darin sind bis zu eine Million Ergebnisse pro Monat enthalten.

3.2.2 Webdesign

In diesem Unterkapitel wird das Webdesign mit seinen Funktionen kurz beschrieben.

3.2.2.1 Loginseite

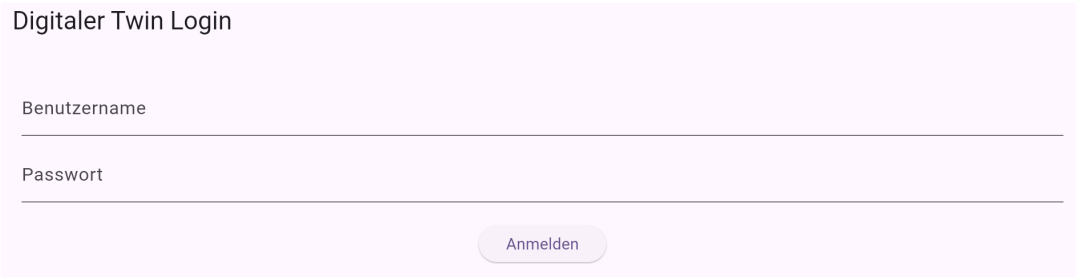

ABBILDUNG 3.2: Login

Dies ist die Login-Seite. Sie ist relativ einfach aufgebaut, was sie aber intuitiv verständlich macht. Momentan gibt es beim Digital Twin nur ein Login mit einem hardcodierten Passwort, um die kostenpflichtigen API-Zugänge von Dehashed und Shodan gegen unautorisierten Zugang zu schützen. Nach dem Einloggen gelangt man direkt zur Hauptseite. Nach zehn Minuten oder beim Schliessen des Browsers erfolgt eine automatische Abmeldung.

3.2.2.2 Hauptseite

Bei dieser Dokumentation zur Webseite wurde die Hauptseite aus Gründen der Übersichtlichkeit in ihren Einzelteilen beschrieben.

Oben auf der Hauptseite kann die gesuchte Domain eingegeben werden. Oben rechts kann man sich ausserdem ordnungsgemäss ausloggen.

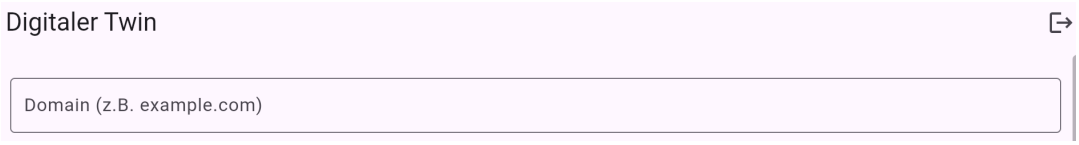

ABBILDUNG 3.3: Auswahl der Domain

Unterhalb der Domainauswahl können mithilfe von Checkboxen die Attribute ausgewählt werden, mit denen interagiert werden soll. Dabei können alle Attribute oder auch nur einige ausgewählt werden. Wenn man auf das Informationssymbol neben der Checkbox klickt, öffnet sich die Infoanzeige.

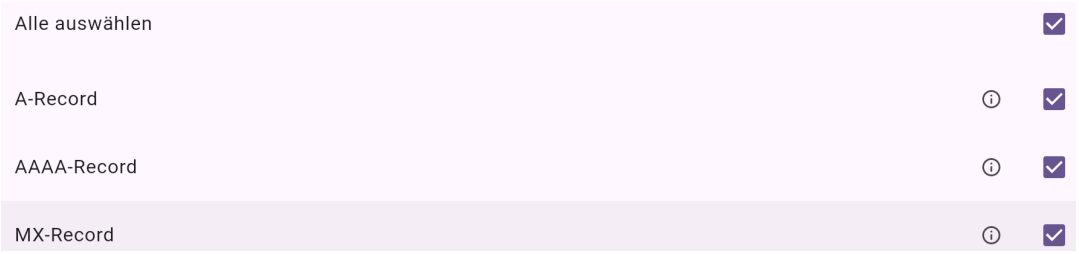

ABBILDUNG 3.4: Auswahl der Attribute

Wenn man auf die Infoanzeige des Attributs „E-Mail" klickt, sieht man gleich eine Beschreibung des Attributs mitsamt den Eigenschaften, die gesammelt werden. Die Spalten entsprechen hier den Eigenschaften, was zugleich mit der späteren Anzeige der Ergebnisse übereinstimmt.

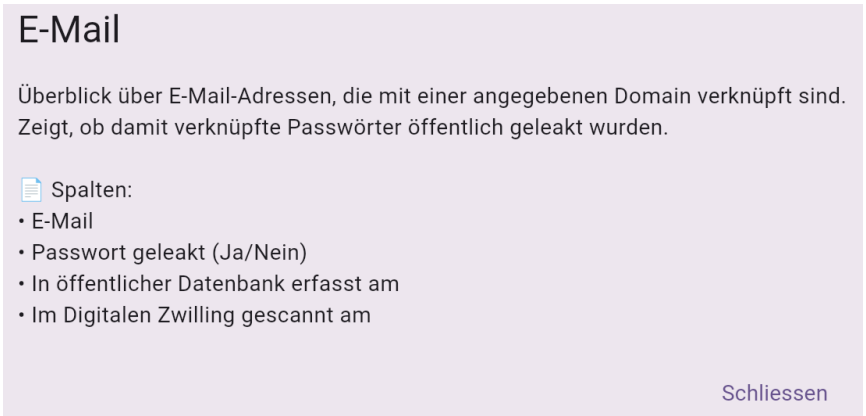

ABBILDUNG 3.5: Infoanzeige des Attributs „E-Mail"

Unterhalb der Auswahl der Attribute befinden sich drei Buttons.
Mit dem ersten Button wird der Scan gestartet und die Daten werden automatisch in der Datenbank der Webseite Digital Twin gespeichert.
Mit dem Anzeige-Button werden die Ergebnisse direkt auf der Webseite angezeigt.
Wenn man die Ergebnisse exportieren möchte, kann man das mit dem letzten Button *Exportieren* machen. Dieser exportiert die Ergebnisse automatisch im JSON-Format in den Download-Ordner. Momentan wird die JSON-Datei so formatiert, dass sie sehr gut lesbar ist. Dies erfordert jedoch mehr Zeilen und somit auch mehr Speicherplatz. Wenn das nicht erwünscht ist, kann in der Klasse *home_screen* im Ordner *Fronted/lib/screens* der Parameter *bool pretty* der Methode
ApiService.exportJson von *True* auf *False* gesetzt werden, um Speicherplatz zu sparen.

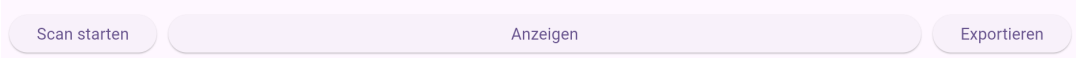

ABBILDUNG 3.6: Buttons und ihre Funktionen

Wenn man auf den Button *Anzeigen* klickt, werden die Ergebnisse der ausgewählten Attribute angezeigt. Als Beispiel zeigen wir hier die Ergebnisse des E-Mail-Scans der Domain apb-architekten.ch. Dort ist auf einen Blick zu erkennen, dass das Passwort der E-Mail-Adresse ra@apb-architekten.ch im Klartext veröffentlicht wurde. Diese E-Mail wurde im Jahr 2015 in der öffentlichen Datenbank der Wayback Machine gespeichert. Nichtsdestotrotz kann diese E-Mail mit diesem Passwort in diversen Onlinekonten aktiv sein.
Wenn ein Attribut mehr als zehn Einträge hat, wird eine neue Seite veröffentlicht und man kann mit den Buttons *Zurück* und *Weiter* zwischen den Seiten wechseln.

ABBILDUNG 3.7: Ergebnisse

3.2.3 Datenbank

Als Datenbank wurde PostgreSQL verwendet, das automatisch mit dem Backend über den Docker Container gestartet wird. PostgreSQL ist eine leistungsstarke, objektrelationale Open-Source-Datenbank. [42]
Jedes Attribut hat eine eigene Tabelle. Für jedes Attribut ist jede Tabelle gleich aufgebaut. In der ersten Spalte steht immer die Domain, in der zweiten das JSON-File mit den gescannten Daten dieses Attributs bezüglich dieser Domain. Bei jedem neuen Scan der Domain wird der alte Eintrag gelöscht und durch den neuen ersetzt.
Hier ist noch ein Beispiel für die Einträge der Tabelle „A-Record".

Domain	Json-File
zhaw.ch	`{"A": [{"IPv4-Adresse": "160.85.192.83", "Im ...}]}`
3-plan.ch	`{"A": [{"IPv4-Adresse": "149.126.4.88", "Im ...}]}`
amag.ch	`{"A": [{"IPv4-Adresse": "51.105.166.17", "Im ...}]}`

TABELLE 3.26: Einträge der Tabelle „A-Record"

Kapitel 4

Resultate

In diesem Kapitel werden die nackten Zahlen der Applikation vorgestellt. Insgesamt wurden 384 Firmen aus dem Raum Winterthur ausgewählt und sämtliche Attribute analysiert. [43] Die Anzahl der Treffer pro Attribut sind in einem Säulendiagramm dargestellt. Auf der Y-Achse sind die Anzahl der Firmen und X-Achse die Anzahl der Ergebnisse pro Attribut abgebildet. Firmen stehen repräsentativ für Domains.
Im Folgenden werden die wichtigsten Attribute näher erläutert; die übrigen können bei Interesse im Anhang nachgelesen werden. Zusätzlich wurden zwei Firmen kontaktiert, damit die gesammelten Daten auf ihre Qualität geprüft werden können. Somit besteht die Möglichkeit, die Ergebnisse auf ihre Quantität und Qualität zu bewerten.

Hinweise zur Interpretation der Resultate sowie eine Einordnung der Zahlen erfolgen im nächsten Kapitel, das sich mit Diskussion und dem Ausblick befasst.

4.1 Analyse der Attritbute

In diesem Kapitel werden nur die Ergebnisse der auffälligen Attribute analysiert. Die unauffälligen werden der Vollständigkeit halber im Anhang beschrieben.

4.1.1 A-Records - Verteilung

Die Grafik zeigt deutlich, dass die überwiegende Mehrheit der Firmen, IPv4-Adressen verwendet. Nur bei fünf Firmen hat die Applikation keine Adresse gefunden und bei 47 Firmen hat man sogar mehr als zwei Adressen gefunden. Bei 384 Firmen hat man bei 86.45% genau eine einzige IPv4-Adresse gefunden.

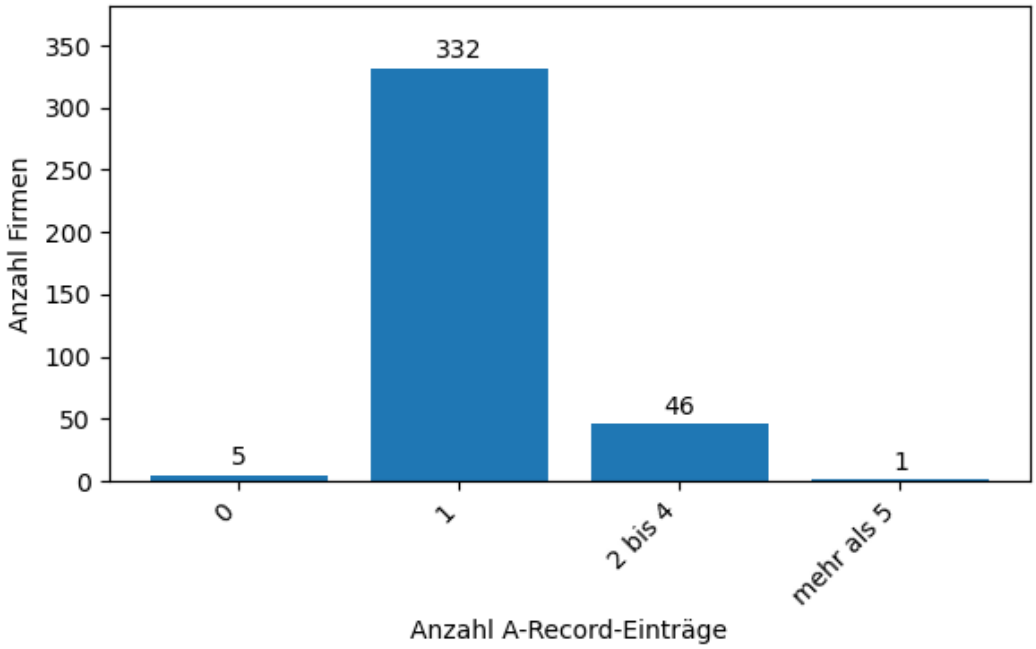

ABBILDUNG 4.1: A-Records - Verteilung pro Firma

Bei nicht gefundenen Adressen muss eine Überprüfung stattfinden, weil diese im Normalfall existieren müssten. Bei den fünf nicht gefunden Adressen sind die folgenden Gründe ausschlaggebend:

- Zwei Domains existieren nicht mehr, weil die Firmen aufgelöst wurden.
- Theoretisch existiert die Domain nicht für die Öffentlichkeit, aber es existiert ein SOA-Eintrag.
- Eine Domain ist falsch konfiguriert. Beispiel: Anstelle von example.com würde nur www.example.com funktionieren. Dieses Problem könnte man lösen, indem man zusätzlich auch nach www.example.com scannt.[44]
- Nach manueller Überprüfung existieren die IPv4- und IPv6-Adressen, aber diese sind aber in unserem Resultat als inexistent verzeichnet. Wahrscheinlich gab es Fehler während des Analysevorgangs. Ein zweiter Versuch der Analyse hätte das Problem behoben.

Typischerweise hat jede Domain eine IPv4-Adresse, weil diese weit verbreitet ist. Es ist sehr unüblich, wenn eine Domain keine IPv4-Adresse besitzt und diese im Netz erreichbar sein soll. Eine Ausnahme wäre, wenn die Domain eine IPv6-Adresse, aber keine IPv4-Adresse hat. Dies ist aber unüblich, weil IPv6 nicht weit verbreitet ist. [45] Viele Domains haben entweder nur IPv4 oder IPv4 und IPv6.

4.1.2 Zertifikate - Verteilung

Bei vier Firmen wurden weniger als zehn Zertifikate gefunden. Bei 90% der analysierten Firmen wurden mehr als zehn Zertifikate gefunden und bei über 60% sogar mehr als 100 Zertifikate. Die meisten Zertifikate sind veraltet und nur wenige sind aktuell. Dies erklärt die hohe Anzahl. Interessanterweise hat die Applikation bei 34 Firmen (9%) kein einziges Zertifikat gefunden, was nicht möglich sein sollte, weil jede seriöse Webseite Zertifikate hat.

Auf der Webseite von crt.sh (Tool zur Nutzung externer Zertifikatdatenbanken) kann bei manueller Eingabe der Domains beobachtet werden, dass Zertifikate vorhanden sind. Es liegt nahe, dass die Applikation Probleme mit der Verbindung zum crt.sh-Tool hatte und somit keine Zertifikate speichern konnte. Wenn kein Zertifikat gefunden wurde, dann kann dies mit einem zweiten automatischen Scan behoben werden.

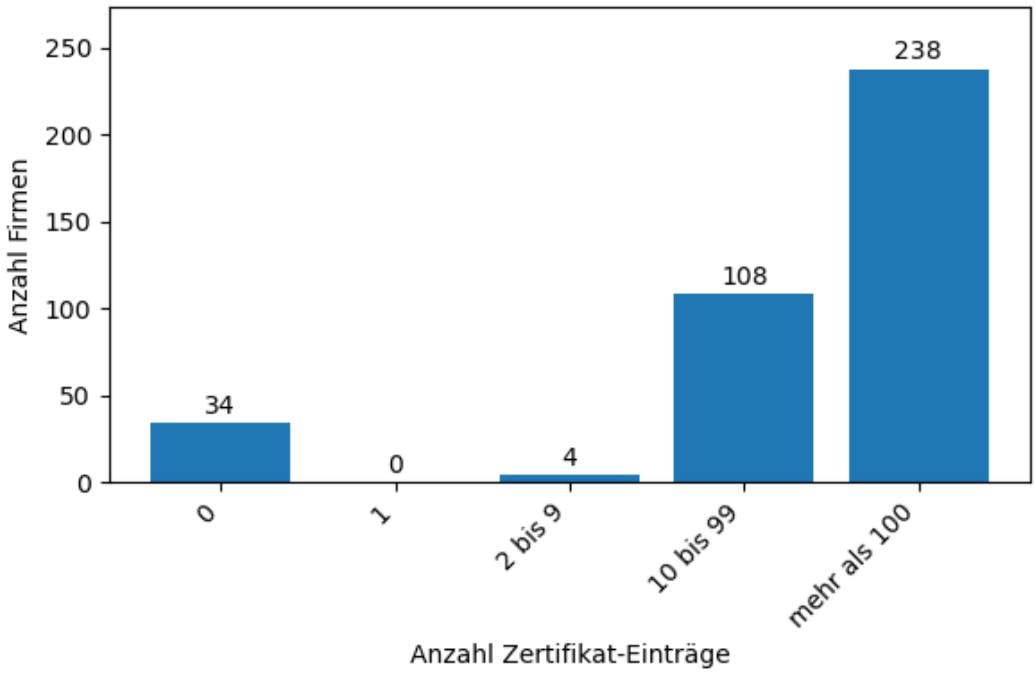

ABBILDUNG 4.2: Zertifikate - Verteilung pro Firma

4.1.3 MX-Records - Verteilung

Die Grafik zeigt die Anzahl gefundener Mailserver pro Domain. Bei 239 Domains wurde genau ein Mailserver gefunden. Bei 124 Domains wurde mindestens zwei Mailserver gefunden und bei 15 sogar über fünf Mailserver. Das Interessante ist nur, dass bei sechs Domains kein einziger Mailserver gefunden wurde. Die Überprüfung der Daten hat folgendes ergeben:

- Zwei Domains existieren nicht mehr.
- Eine Domain leitet auf eine andere weiter, da die Firma hinter der weiterleitenden Domain mit der Firma der empfangenden Domain fusioniert hat. Die weiterleitende Domain verfügt daher über keinen eigenen Mailserver, während die empfangende Domain über eine vollständige Mailinfrastruktur verfügt.
- Eine Domain ist ein Subdomain und die Mailserver-Einträge im DNS findet man wiederum im Hauptdomain.
- Zwei Firmen haben nach manueller Analyse keine Mailserver. Eine dieser Firmen hat nach einer persönlichen Abklärung den Verdacht bestätigt, dass sie keine Mailserver verwenden.

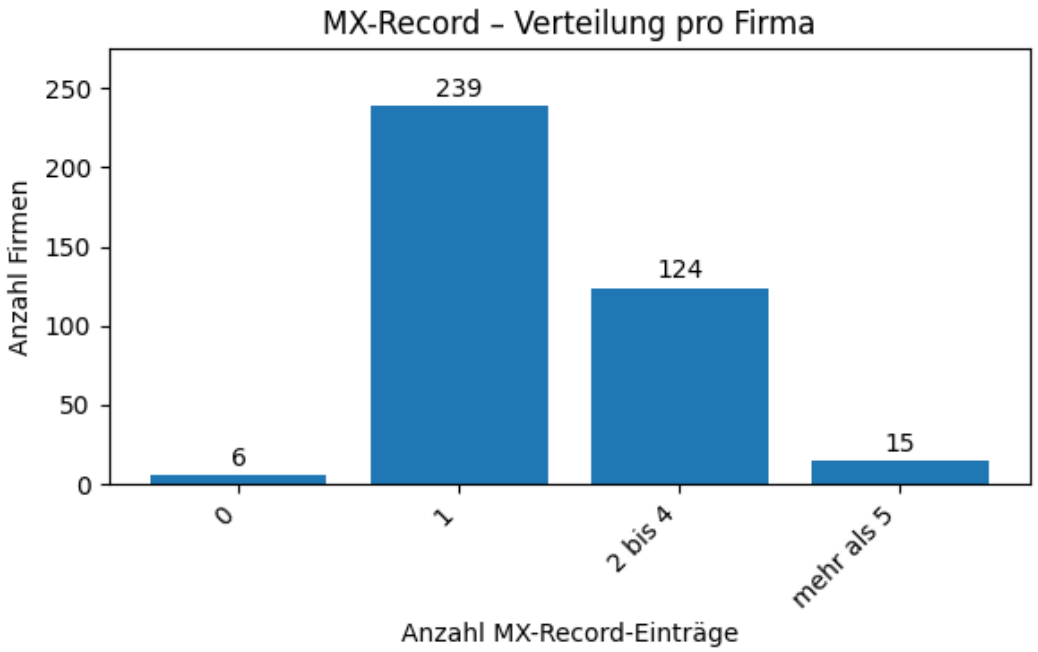

ABBILDUNG 4.3: MX-Records - Verteilung pro Firma

4.1.4 E-Mails – Passwort Leak

In dieser Grafik wurde analysiert, bei wie vielen Firmen bereits das Passwort von E-Mail-Adressen geleakt wurden. Die Anzahl der analysierten Firmen beläuft sich auf 171. Die geringe Anzahl an Testdaten lässt sich damit erklären, dass das Tool zur Leak-Analyse (dehashed) kostenpflichtig ist und die Kosten bei grösseren Datenmengen stark ansteigen. Bei 43 Firmen (25.15%) wurde von mindestens einer E-Mail-Adresse das Passwort geleakt. Diese Grafik wurde in der ersten Testphase erstellt und ist nur bedingt aussagekräftig. Unter den 43 Mailadressen, die geleakt wurden, sind jene dabei, die nicht mit der Domain der Firma enden, aber auf der Webseite gefunden wurden. Diese falschen E-Mail-Adressen könnten Dummy-Adressen sein oder im Impressum auf den Webseitenersteller verweisen.

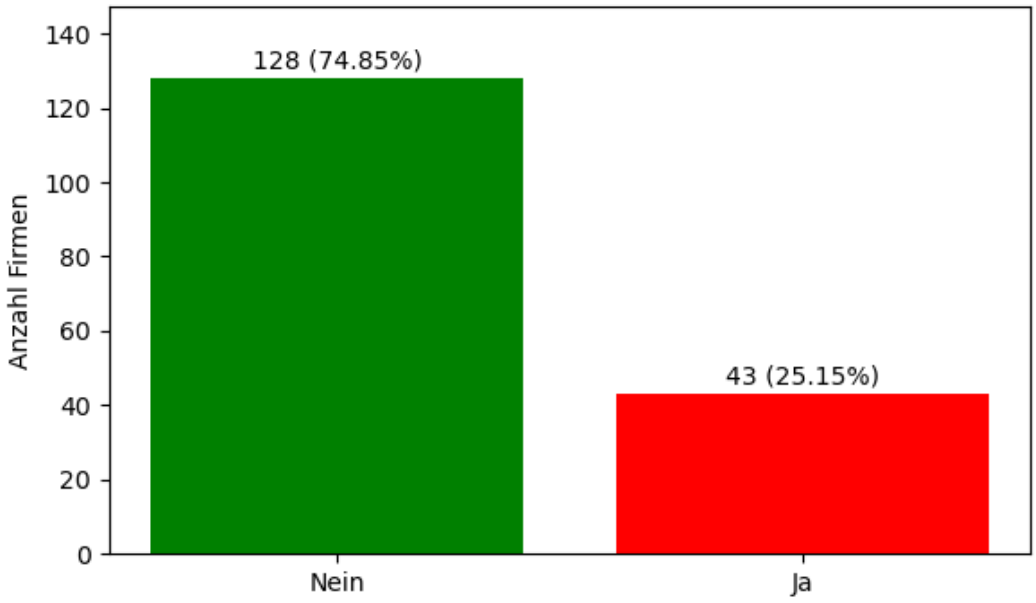

ABBILDUNG 4.4: E-Mail - Passwort geleakt? - Unmodifiziert

In dieser Grafik haben wir die richtigen Daten. Es wurden nur die E-Mail-Adressen als geleakt markiert, die auch den Firmen selbst gehören. Dabei sind neu 31 statt 43 Firmen betroffen. Statt jede vierte ist jetzt nur jede fünfte Firma betroffen. Der Nachteil dieser Methode ist jedoch, dass es, wie beim MX-Record beschrieben, insbesondere Kleinunternehmer gibt, die beispielsweise eine Gmail-Adresse als Geschäftsadresse verwenden. Diese E-Mails würden mit der neuen Methode ignoriert werden.

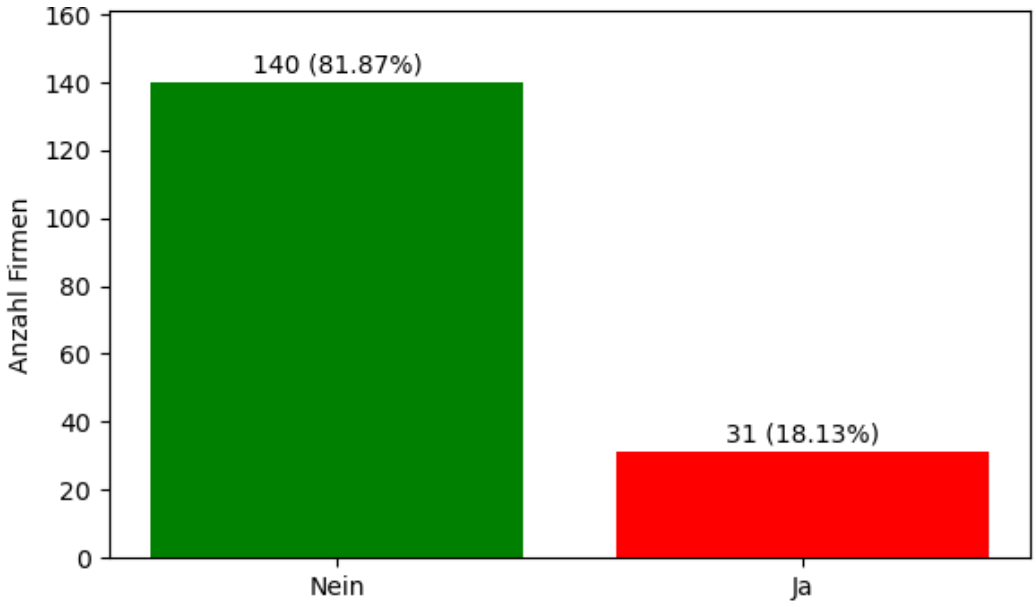

ABBILDUNG 4.5: E-Mail - Passwort geleakt? - Modifiziert

4.1.5 Telefonnummern - Passwort Leak

In dieser Grafik wurde analysiert, ob bei den analysierten Firmen Konten existieren, bei denen sowohl Telefonnummern als auch Passwörter geleakt wurden. Bei 185 untersuchten Firmen sind 11 (5.95%) betroffen. Dies bedeutet, dass bei diesen Firmen mindestens ein Konto erfasst wurde, bei dem sowohl eine Telefonnummer als auch ein Passwort öffentlich bekannt ist. Die Kombination stellt ein erhebliches Sicherheitsrisiko dar und macht diese Firmen angreifbar für Social Engineering.

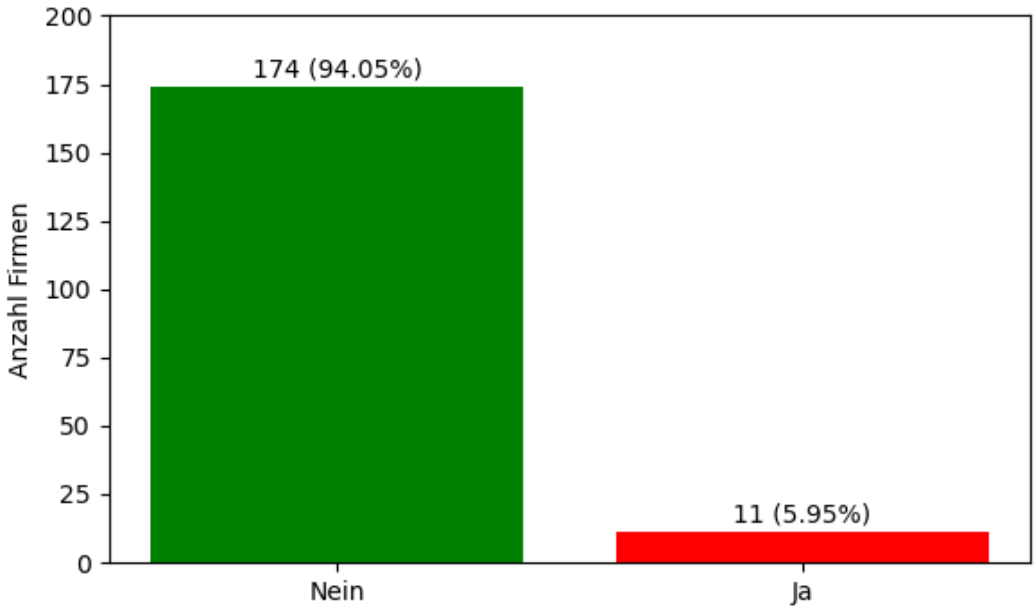

ABBILDUNG 4.6: Telefonnummer - Passwort geleakt?

4.1.6 NS-Records - Verteilung

Diese Grafik zeigt, dass über 98% der analysierten Firmen mehr als zwei Nameserver haben. Nur bei fünf Firmen wurde kein einziger Nameserver gefunden, was nicht der Normalität entspricht.

- Von diesen fünf Firmen existieren zwei nicht mehr.
- Bei einer Domain handelt es sich um eine Subdomain, bei der sich die Einträge in der Hauptdomain befinden.
- Bei zwei Domains konnten bei manueller Überprüfung je drei Nameserver gefunden werden. Das führt zu einer Fehleranfälligkeit von 0.5%. Der Digitale Twin könnte um die folgende Funktion ergänzt werden: Wenn er beim Scannen keinen NS-Eintrag erhält, scannt er automatisch ein zweites Mal.

Grundsätzlich sind mindestens zwei Nameserver empfohlen und jede Domain braucht einen Nameserver. Das bedeutet, wenn eine Domain bei der Analyse angeblich null Nameserver hat, dann muss das überprüft werden. Sogar wenn nur ein Nameserver existiert, ist das ungewöhnlich. [46]

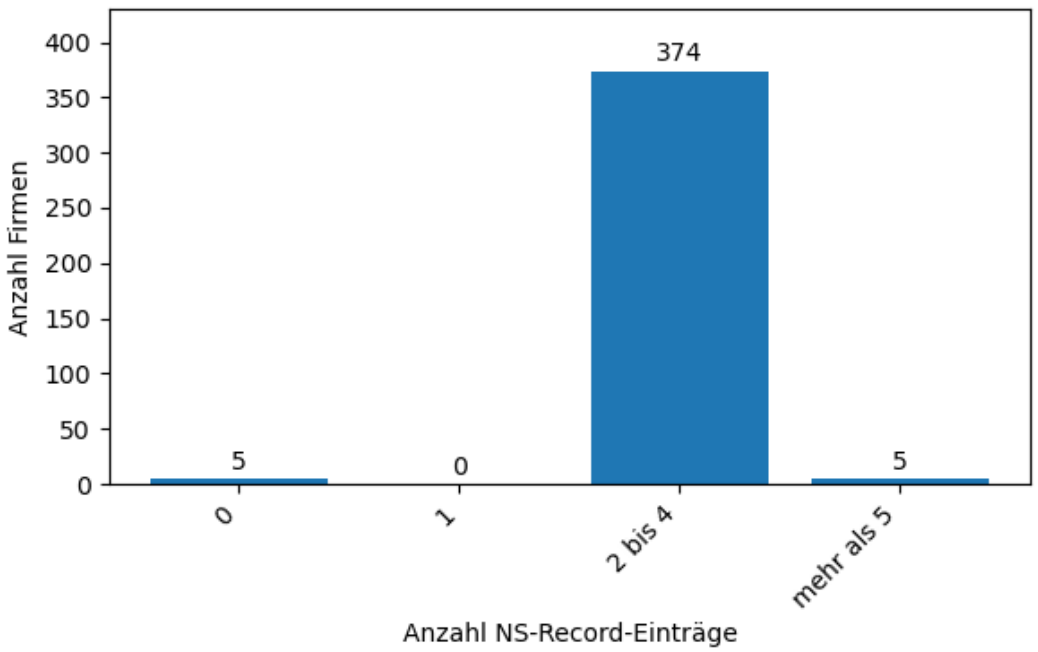

ABBILDUNG 4.7: NS-Records - Verteilung pro Firma

4.1.7 Dienste - Verteilung und CVE

Diese Grafik zeigt die Verteilung von Dienst-Einträgen. Die grosse Mehrheit der Firmen (281) verfügt über keinen Eintrag. 36 Firmen haben genau einen Dienst-Eintrag, 38 Firmen bis zu neun und 29 Firmen mehr als zehn. Diese Verteilung ist nicht ungewöhnlich.

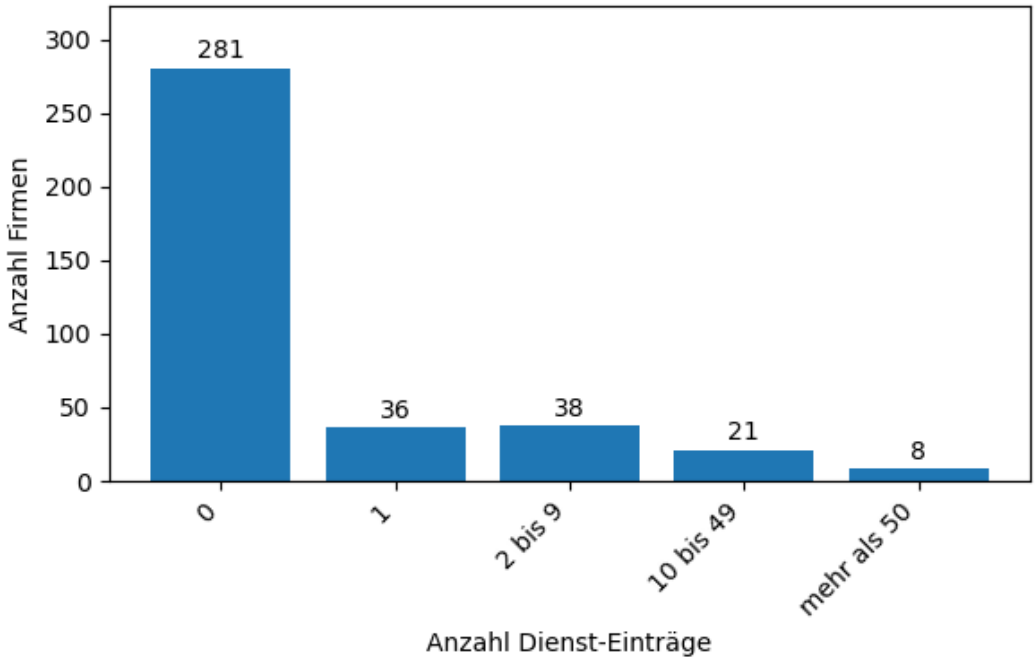

ABBILDUNG 4.8: Dienste - Verteilung pro Firma

Bedenklich ist hingegen der CVE-Wert. Die Grafik zeigt, ob bei einer Firma eine bekannte Sicherheitslücke in einem identifizierten Dienst vorliegt. Bei 28 Firmen (7.29%) wurde mindestens eine solche Schwachstelle festgestellt. Da ein CVE nur Schwachstellen in Bezug auf bestimmte Softwareversionen anzeigt, kann eine Firma diese durch Konfiguration gesichert haben, ohne dass dies im CVE berücksichtigt wird. Somit würde nur das CVE-System als Schwachstelle angezeigt werden, obwohl dies nicht mehr zutrifft. Zudem hat Shodan nicht immer Zugriff auf aktuelle Daten, sodass diese Schwachstellen möglicherweise bereits behoben sind, weil die Firma die Software auf eine sichere Version aktualisiert hat.

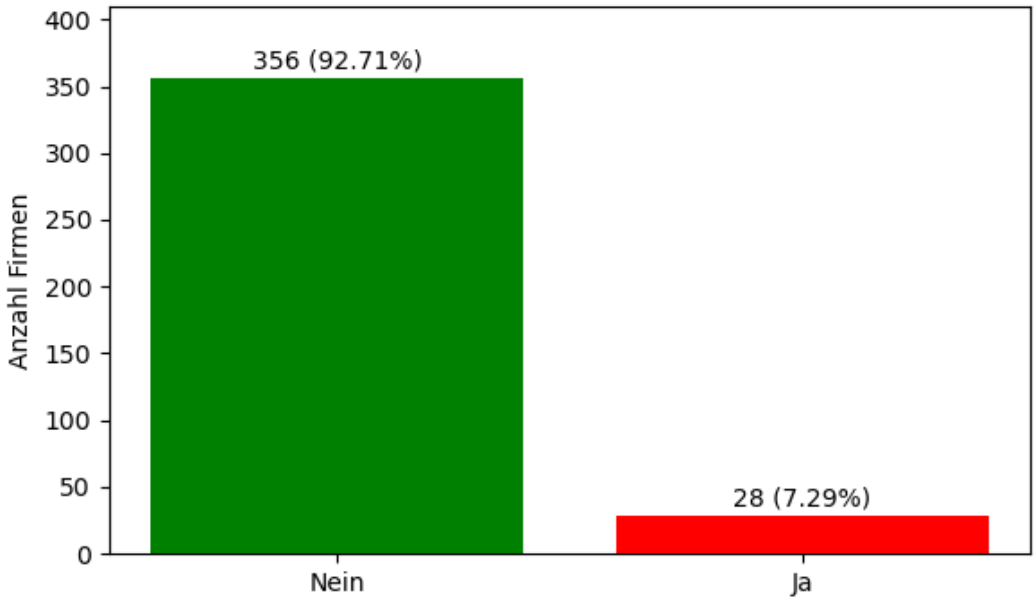

ABBILDUNG 4.9: CVE gefunden? - Verteilung pro Firma

4.1.8 SOA-Records - Verteilung

Die SOA-Verteilung zeigt, dass unter 384 Firmen bei dreien keine SOA-Einträge gefunden wurden, was definitiv nicht der Normalität entspricht.[47] Die Analyse hat folgendes ergeben:

- Zwei Domains existieren nicht mehr und liefern daher kein korrektes Ergebnis.
- Bei einer Domain handelt sich um eine Subdomain. Die DNS-Einträge befinden sich in der Hauptdomain.
- Interessanterweise gibt es ein Domain, welches keine IPv4- und IPv6-Adressen hat, aber einen SOA-Eintrag. Wahrscheinlich wird diese Domain intern genutzt und ist nicht für die Öffentlichkeit bestimmt.

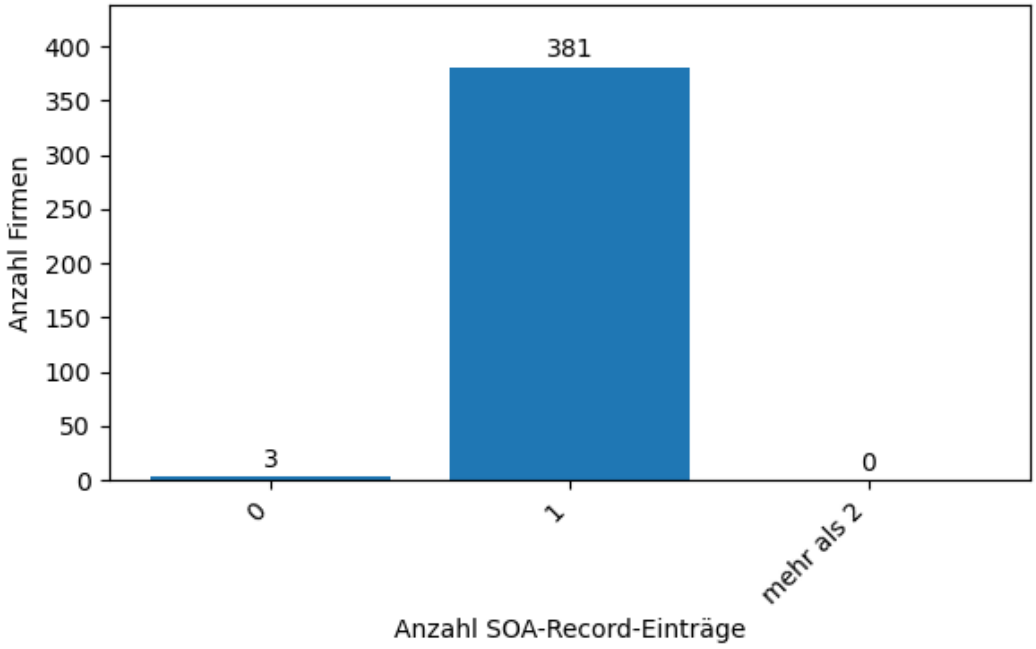

ABBILDUNG 4.10: SOA-Records - Verteilung pro Firma

4.2 Feedback zweier Firmen

Um die erhaltenen Daten auf ihre Qualität und Richtigkeit zu überprüfen, wurde in diesem Zusammenhang mit zwei Firmen zusammengearbeitet. Eine davon ist ein IT-Dienstleister und die andere ist im Bereich der Elektroplanung tätig.

Bei der Analyse der Daten des IT-Dienstleisters stimmen die IPv4- und IPv6-Adressen überein. Sie nutzen die Mail- und Nameserver der Abraxas Informatik AG. Alle aktuellen E-Mail-Adressen und Telefonnummern wurden ebenfalls gefunden und sind korrekt.
Interessanterweise konnte die Applikation Subdomains herausfinden, die nicht für die Öffentlichkeit bestimmt sind. Für die Suche nach Subdomains wurde das Tool „sublist3r" mit „crt.sh" verwendet. Diese Subdomains waren früher öffentlich zugänglich, weshalb entsprechende Zertifikate erstellt wurden. Die Zertifikate sind inzwischen abgelaufen und die Subdomains nicht mehr öffentlich zugänglich. Da crt.sh unzählige abgelaufene Zertifikate gespeichert hat, kann die Existenz nicht mehr öffentlicher Subdomains nachgewiesen werden. Eine nicht öffentlich zugängliche Subdomain, die intern jedoch noch benutzt wird, stellt eine Angriffsfläche dar, die von Angreiferinnen und Angreifer ausgenutzt werden könnte. Deshalb sind diese Informationen für eine Firma besonders interessant. Sie müssen zusätzlich abgesichert werden. Firmen, die glauben, ihre Subdomain sei geheim, sichern diese weniger ab. Dadurch kann sie für Angreiferinnen und Angreifer zu einem Angriffspunkt werden. In diesem Fall existierte die Subdomain nach der Abklärung mit der zuständigen Person des IT-Dienstleisters nicht mehr und stellt somit kein Sicherheitsrisiko für die Firma dar.

Bei der Analyse der Elektroplaner-Firma stimmen die IPv4- und IPv6-Adressen. Sie haben einen Outlook-Mailserver und ns1.cyon.ch als Nameserver, was ebenfalls korrekt ist. Cyon ist ein Webseitenprovider. [48]
Bei der Analyse kam ausserdem heraus, dass die Applikation nur eine E-Mail gefunden hat, obwohl mehrere E-Mail-Adressen verfügbar waren. Zusätzlich wurde festgestellt, dass die Applikation nur eine E-Mail-Adresse gefunden hat, obwohl mehrere E-Mail-Adressen verfügbar waren. Die Applikation arbeitet mit der Wayback Machine und die Suche wurde aus Performance-Gründen auf die ersten 50 URLs beschränkt. Es besteht die Möglichkeit, dass sich die restlichen E-Mail-Adressen auf der 51. URL befinden.
Zusätzlich haben wir keine Subdomains gefunden. Die Firma hat bestätigt, dass sie keine Subdomains hat. Shodan hat beim Scan-Dienst jedoch zusätzlich eine veraltete Domain gefunden. Dies könnte darauf hindeuten, dass Shodan veraltete Daten besitzt oder dass diese Subdomain ohne das Wissen der Firma doch noch existiert. Da diese Firma selbst über keinen IT-Experten verfügt, ist davon auszugehen, dass sie nicht den vollständigen Überblick über ihr IT-System hat.
Es wurde ebenfalls eine Telefonnummer gefunden. Laut der Firma ist nur diese eine öffentlich einsehbar. Es besteht auch hier die Möglichkeit, dass weitere Direktwahlnummern ohne das Wissen der Firma öffentlich auffindbar sind, die von uns jedoch nicht gefunden wurden.

4.2.1 Fazit

Bei der Analyse der Firmen kam heraus, dass die Daten stimmen. Beide Firmen konnten interessante Erkenntnisse aus ihren Daten gewinnen. Es kann jedoch der Fall eintreten, dass die gefundenen Daten unvollständig sind.
Mit passivem OSINT ist es schwierig, alle Daten zu ermitteln. Beim IT-Dienstleister wurden Subdomains gefunden, die früher einmal öffentlich waren. Beim Elektroplaner konnten nicht alle E-Mail-Adressen gefunden werden, da die Suche aus Gründen der Performance eingeschränkt wurde. Es ist auch möglich, dass ein kleiner Teil der Daten veraltet ist. Vor allem die Zertifikate sind abgelaufen, was darauf zurückzuführen ist, dass crt.sh alle Zertifikate speichert. Ausserdem ist bekannt, dass Shodan alte Daten hat.
Veraltete Daten können jedoch auch nützlich sein. Sie können Angreiferinnen und Angreifer bei Social Engineering helfen. Zudem sind manche Daten zwar veraltet und nicht öffentlich, aber immer noch in Betrieb. Bei diesen Angriffsflächen wird die Sicherheit in der Firma weniger beachtet, weshalb sie eine zusätzliche Gefahr darstellen. Manchmal reicht nämlich ein unbedeutender Angriffspunkt, um einer Firma grossen Schaden zuzufügen.
Trotz dieser Herausforderungen wurde ein sehr gutes Ergebnis erzielt.

Kapitel 5

Diskussion und Ausblick

5.1 Einordnung der Ergebnisse

Die Applikation ist ein Prototyp für die Sammlung von Informationen eines Unternehmens bzw. anhand der Information eines Digital Twins eines Unternehmens. Dieses beinhaltet das passive OSINT und bildet eine gute Lösung für KMUs. [11]

5.2 Relevanz für KMU

Die überwiegende Mehrheit der Schweizer Bevölkerung arbeitet in einer KMU, denn diese machen 99% der Schweizer Firmen aus. [1] Viele dieser Firmen haben leider keine ausreichenden Expertise in der IT, weil sie sparsam und mit Bedacht ihre Ressourcen verwenden müssen. Jede Firma in der heutigen Zeit benutzt das Internet und gibt dabei ihre Daten preis. Da geht schnell die Übersicht verloren, welche Daten überhaupt öffentlich einsehbar sind. Hier kann es schnell passieren, dass Daten öffentlich sind, die es nicht sein sollten. Unsere Applikation ist eine Bereicherung für die Datensicherheit dieser Firmen, weil sie mit nur einer einfachen Eingabe der Domain ihrer Webseite alle aktuellen und ehemaligen öffentlichen Daten ihrer Firma auf einer Seite übersichtlich zusammengefasst einsehen können. Der Grundsatz, dass der Verteidiger all seine Infrastruktur schützen muss, aber der Angreifer nur eine Schwachstelle finden muss, macht es unerlässlich, eine Übersicht über seine Daten zu haben, die Spuren in der Öffentlichkeit hinterlassen haben. So kann der Verteidiger (KMU) das Wissen erlangen, welche Daten jetzt besonders geschützt werden müssen.

5.3 Erreichung der Ziele

Mit dieser Arbeit wurde auch ein wichtiger Schritt in die Richtung der Verbreitung von Digital Twins im Bereich der Cyber Security gemacht. Digital Twins existieren meistens in Form von physiche Sachen Digital abzubilden. Diese Arbeit behandelt die Errichtung eines Digital Twins einer Webseite bzw. einer virtuellen Technologie, was in der Praxis selten existiert. Mit dieser Arbeit wurde gezeigt, dass es nicht nur möglich, sondern auch sinnvoll ist, virtuelle Technologien digital abzubilden. Die im Kapitel 1.2.2 formulierten Ziele wurden alle erreicht. Die Applikation ist modular aufgebaut und dadurch ist es möglich, neue Attribute einfach zu implementieren. Es ist Passiv OSINT Konform, weil die Applikation nur auf öffentliche Daten zugreift und auf einer Webseite schön und übersichtlich darstellt.

Stand jetzt ist es wichtig zu erwähnen, dass der digitale Twin, gut aber nicht perfekt ist. Die Richtigkeit der Daten kann nach diversen Tests gewährleistet werden, aber es besteht die Möglichkeit, dass die Daten unvollständig sind. Mit passivem OSINT können gute und viele Ergebnisse gewonnen werden. Da wir aufgrund des passiven OSINT nur auf öffentliche Daten zugreifen konnten, kann die vollständige Abdeckung der Daten nie sichergestellt werden. Es kann bei allen analysierten Attributen vorkommen, dass der Scan aufgrund eines kurzen Verbindungsfehlers kein Resultat liefert. In der Regel wird beim zweiten Versuch ein Ergebnis generiert. Die Verbindungsfehler sind externer Natur und können nicht beeinflusst werden. Derzeit muss der zweite Scan manuell gestartet werden. In Zukunft sollte im Backend überprüft werden, ob der Scan ein leeres Ergebnis liefert. Falls ja, dann sollte der Scan automatisch wiederholt werden. So kann sichergestellt werden, ob tatsächlich keine Daten im Netz vorhanden sind oder ein Verbindungsfehler die Ursache war.

5.4 Technische Herrausforderungen und Lösungen

5.4.1 Wayback Machine

Die Suche nach E-Mail-Adressen und Telefonnummern mithilfe der Wayback Machine kann sehr zeitaufwendig sein. Das liegt daran, dass die Wayback Machine automatisierte Suchanfragen sperrt. Um dies zu umgehen, wurde nach jedem Scan eine zufällige Pause von fünf bis zwölf Sekunden eingebaut. Dadurch hat sich die Scanzeit allerdings stark erhöht. Deshalb wurde die Suche auf 50 URLs begrenzt. Somit konnte die Scandauer gesenkt werden, allerdings kann die Qualität und Quantität der Daten dadurch nicht mehr zu 100% gewährleistet werden.

5.4.2 Crt.sh-Zertifikatssuche

Die Erfahrung mit dem Tool Crt.sh, das für die Ermittlung der Zertifikate verwendet wird, hat gezeigt, dass bei einer zu grossen Anzahl an Zertifikaten das Ergebnis automatisch von Crt.sh verkleinert wird und nicht alle Zertifikate angezeigt werden. Es ist ratsam, die URL, die für die Suche verwendet wird, in Zukunft anzupassen. Neben der aktuell genutzten URL „https://crt.sh/?q=example.com“ sollte auch die URL „https://crt.sh/?q=example.com&exclude=expired&group=none“ verwendet werden. Mit dem Zusatz „&exclude=expired&group=none“ werden nur die aktuellen Zertifikate angezeigt. Durch diese Anpassung wird sichergestellt, dass die aktuellen Zertifikate sicher angezeigt werden, aber auch möglichst viele der abgelaufenen Zertifikate, die immer noch nützliche Informationen über das Unternehmen liefern.

5.5 Limitierung des Digital Twin

Der modulare Aufbau macht es möglich, auf einfache Art und Weise Attribute hinzuzufügen und zu implementieren. Nachteilig ist, dass der modulare Aufbau das Interface der Attribute einschränkt. Die Visualisierung der Attribute ist überall gleich aufgebaut und kann nicht spezifiziert werden. Manche Features können für gewisse Attribute nutzvoll sein, was infolge der Implementierung durch den modularen Aufbau nicht möglich ist. Ausserdem lag der Fokus in dieser Arbeit mehr auf der korrekten Funktionsweise der Applikation und weniger auf der Anzahl der Attribute. In Zukunft sollte der Fokus auf der Implementation neuer Attribute liegen.

5.6 Erweiterung um Aktives OSINT

Wie erwähnt beinhaltet diese Applikation nur das passive OSINT. Ein Must-Have ist zusätzlich das aktive OSINT. Dieses kann die Information direkt beim Ziel suchen und sammeln.[11]
Dies darf aber nur mit dem Einverständnis des Ziels erfolgen. Um dies Sicherzustellen, dürfen nur verifizierte Unternehmen für den Eigengebrauch das akitve OSINT verwenden. Dafür ist eine Login-Funktion notwendig.

5.7 Login- und Sicherheitserweiterung

Derzeit bestehen die Zugangsdaten aus einem Standard-Benutzernamen und einem hardcodierten Passwort. Die Loginfunktion soll weiter ausgebaut werden, wobei der Sicherheitsfaktor mit der Salt-&-Pepper-Technik verstärkt werden soll. Dazu sollen Hash-Funktionen verwendet und eine nach aktuellen Cybersecurity-Standards gesicherte Datenbank genutzt werden.

5.8 Testautomatisierung und Continous Integration

Was Stand jetzt fehlt, sind automatisierte Tests. Leider haben uns hier die Zeit und die nötige Erfahrung gefehlt, für Docker, Python und Flutter gute Unit- und Integrationtests zu schreiben, um die Qualität des Codes und das Zusammenspiel der Technologien automatisch sicherzustellen. In naher Zukunft ist es wichtig, diese Tests zu implementieren. Wenn diese gemacht wurden, dann ist es in ferner Zukunft ratsam, sich dem Thema „Continuous Integration“ zu widmen. Diese Methode ist ein Teilbereich des Testing eines Softwareentwicklungprozesses, wo ein gemeinsames Coderepository erstellt wird. Wenn Codeänderungen in dieses Repository durchgeführt werden, dann startet ein automatisierter Build-Prozess. In diesem Prozess werden alle Tests ausgeführt und die Ergebnisse evaluiert. Es gibt anschliessend ein Feedback an die Entwickelnden, die wenn nötig reagieren können. Dies verbessert die Codequalität, ermöglicht schnelleres Feedback und spart enorm viel Zeit.

5.9 Fazit

Diese Arbeit hat gezeigt, dass sich ein Digital Twin nicht nur für physische, sondern auch für virtuelle Objekte wie eine Unternehmenswebseite mit ihrer Infrastruktur erfolgreich und sinnvoll umsetzen lässt.
Die Webseite „Digitaler Twin” ist ein funktionierender Prototyp, der mithilfe von passivem OSINT öffentlich verfügbare Daten eines Unternehmens sammelt, strukturiert darstellt und zugleich modular erweiterbar ist.
Der Digitale Twin ermöglicht insbesondere KMU einen einfachen Überblick über ihre öffentlich einsehbaren Daten. Daraus können sie ableiten, welche Bereiche ihrer Infrastruktur sie besser schützen müssen.
In Zukunft sollten aktive OSINT, automatisierte Tests und eine sichere Login-Funktion hinzugefügt werden, damit bessere Daten geliefert werden können.

Anhang A

Anhang

A.1 Deklaration zur Nutzung von Künstlicher Intelligenz

Zur Optimierung der vorliegenden Arbeit wurden diverse künstliche Intelligenzen herangezogen.

So wurde der Schreibassistent DeepL Write [49] genutzt, um die Entfernung von Rechtschreibfehlern zu automatisieren.

Es wurde der Übersetzungsassistent DeepL Translate für die englische Übersetzung der Zusammenfassung verwendet.[50]

Zudem wurde das Large Language Model ChatGPT [51] von OpenAI eingesetzt, um einen schnellen Überblick über Texte zu erlangen, komplexe Sachverhalte zu verstehen und die Erstellung von Codes zu unterstützen.

A.2 Projektmanagement

Im Rahmen der Bachelorarbeit fanden wöchentliche Meetings im Umfang von jeweils einer halben Stunde statt. Anwesend waren jeweils die beiden betreuenden Personen sowie die beiden Autoren. Vor den Meetings wurde von den beiden Autoren jeweils eine aktuelle Agenda erstellt und den betreuenden Personen über diese hier verlinkte Github-Seite zur Verfügung gestellt. Die Agenda ist jeweils gleich aufgebaut und besteht aus drei Abschnitten. Im ersten Abschnitt wird der Stand der Arbeit dargestellt, d. h., die Autoren zeigen, was seit dem letzten Meeting im Rahmen der Bachelorarbeit passiert ist. Der zweite Abschnitt zeigt die nächsten aktuellen Ziele der Arbeit. Im letzten Abschnitt stellen die Autoren Fragen zu aktuellen Themen an die betreuenden Personen.

A.3 Code

Der vollständige Softwarecode dieser Projektarbeit wurde der Forschungsgruppe Information Security der ZHAW über die hier verlinkte GitHub-Seite zur Verfügung gestellt.

A.4 Restliche Resulate

In diesem Unterkapitel werden die übrigen Ergebnisse der Attribute beschrieben, die sich erwartungsgemäss verhalten.

A.4.1 Endpoints - Verteilung

Diese Grafik zeigt die Anzahl der Endpoints pro Domain. Bei allen Firmen wurde mindestens ein Endpoint gefunden, aber die meisten haben hunderte von Endpoints und das ist nicht aussergewöhnlich.

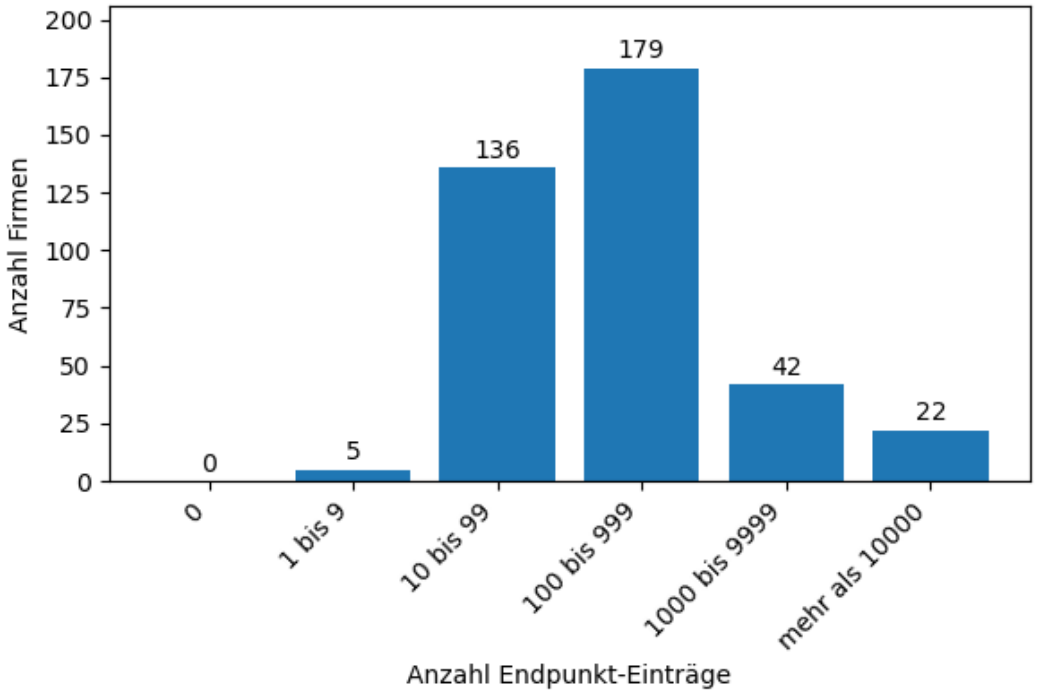

ABBILDUNG A.1: Endpoints - Verteilung pro Firma

A.4.2 E-Mails - Verteilung

Diese Grafik zeigt, wie viele Mailadressen von einer Firma im Netz gefunden wurden. Bei allen 384 Firmen wurde mindestens eine Mail-Adresse gefunden. Bei 130 Firmen (33.85%) wurde genau eine einzige Mailadresse gefunden. Bei 112 Firmen (29.17%) wurden zwei Mailadressen gefunden. Bei 113 Firmen (29.43%) wurden bis zu 20 Mailadressen gefunden. Bei 20 Firmen (5%) wurden sogar mehr als 20 Mailadressen entdeckt. Diese Verteilung verhält sich normal und ist nicht unüblich. Es existieren hier keine Ausreisser.

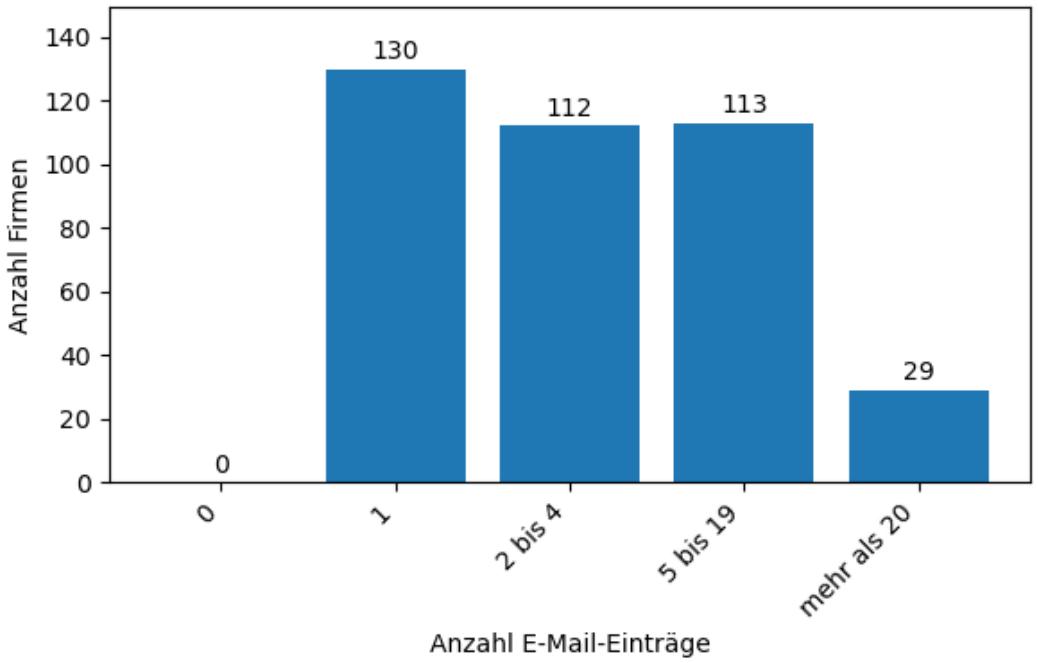

ABBILDUNG A.2: E-Mails - Verteilung pro Firma

A.4.3 AAAA-Records - Verteilung

Es zeigt sich, dass die allermeisten Firmen (33.74%) keine IPv6-Adressen verwendet. Nur 77 Firmen (20%) nutzen eine IPv6-Adresse. Lediglich zehn Firmen verfügen überhaupt mehr als zwei IPv6-Adressen. Diese Verteilung ist nicht unüblich.

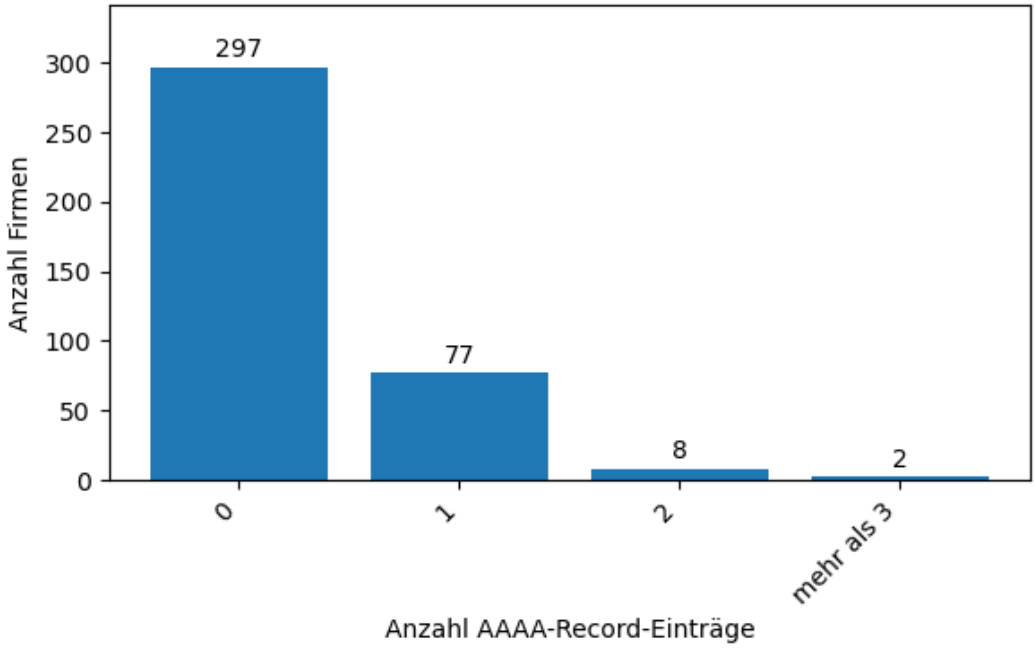

ABBILDUNG A.3: AAAA-Records - Verteilung pro Firma

A.4.4 Telefonnummern - Verteilung

In diesem Säulendiagramm wird die Anzahl der Telefonnummern pro Domain dargestellt. 114 Firmen verfügen über genau eine Firmennummer, 116 Firmen haben bis zu vier Telefonnummern und bei 107 Firmen sind es mehr als fünf. Diese Verteilung ist nicht ungewöhnlich. Manche Firmen veröffentlichen auf ihrer Webseite mehrere Telefonnummern, auch im Zusammenhang mit Stellenausschreibungen werden häufig spezifische Kontaktdaten für die jeweilige Position angegeben.

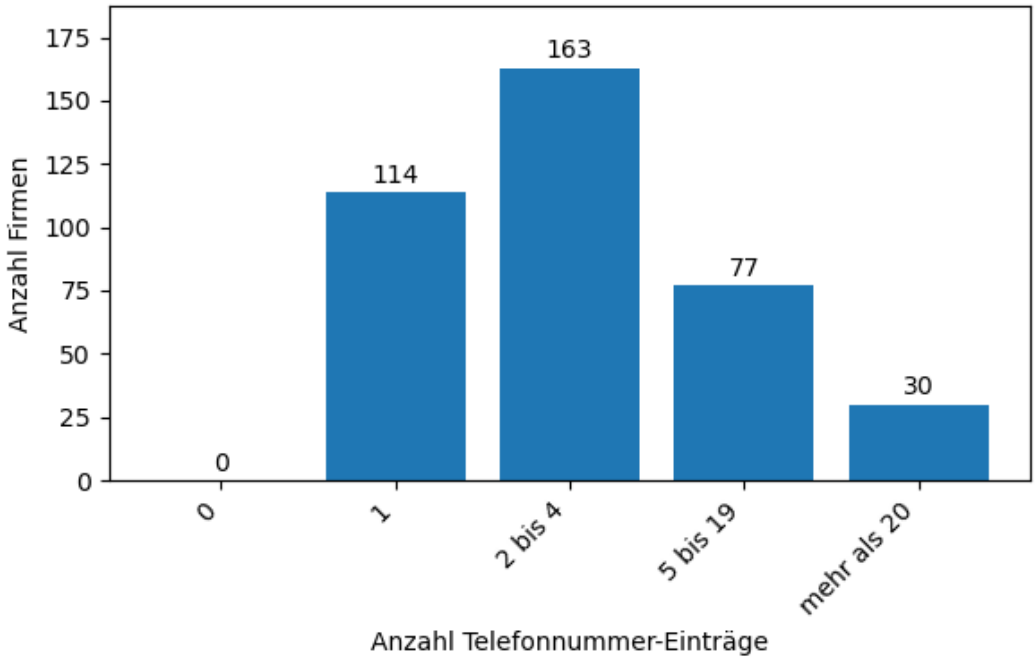

ABBILDUNG A.4: Telefonnummern - Verteilung pro Firma

A.4.5 PTR-Records - Verteilung

Diese Grafik zeigt die Verteilung der PTR-Einträge. Bei 46 Firmen wurde kein PTR-Eintrag gefunden, bei 227 Firmen genau ein Eintrag und bei 111 Firmen mehr als zwei Einträge. Diese Verteilung ist nicht ungewöhnlich.

PTR-Einträge sind für die Reputation einer Firma wichtig, wenn diese eigene Mailserver betreibt. Ohne einen gültigen PTR-Eintrag können versandte E-Mails als Spam eingestuft oder von empfangenden Servern sogar vollständig abgelehnt werden. Da PTR-Einträge nur vom Inhaber der jeweiligen IP-Adresse konfiguriert werden können, was in der Regel der Hosting-Provider und nicht die Firma selbst ist, überrascht es nicht, dass viele Domains keinen eigenen PTR-Eintrag besitzen. Dennoch prüfen viele empfangende Mailserver den PTR-Eintrag, um die Seriosität und die Vertrauenswürdigkeit des Absenders zu bewerten. [52]

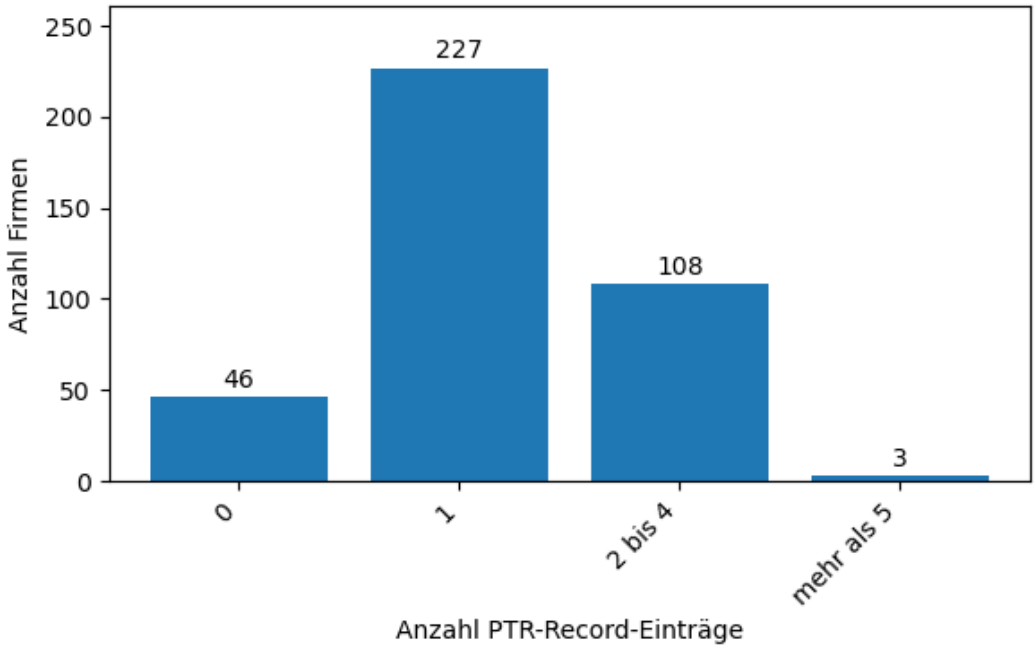

ABBILDUNG A.5: PTR-Records - Verteilung pro Firma

A.4.6 Subdomain-Records - Verteilung

Diese Grafik zeigt die Verteilung von Subdomains. Bei 131 Firmen wurde jeweils genau eine Subdomain gefunden. 94 Firmen verfügen über mehr als zwei, 140 über mehr als fünf und bis zu 49 Subdomains. Nur bei 19 Firmen wurden mehr als 50 Subdomains identifiziert.

Auffällig ist, dass jede analysierte Domain mindestens eine Subdomain aufweist, was eher untypisch ist. Dies lässt sich jedoch dadurch erklären, dass das verwendete Tool sublist3r gelegentlich auch Hauptdomains als Subdomains klassifiziert.

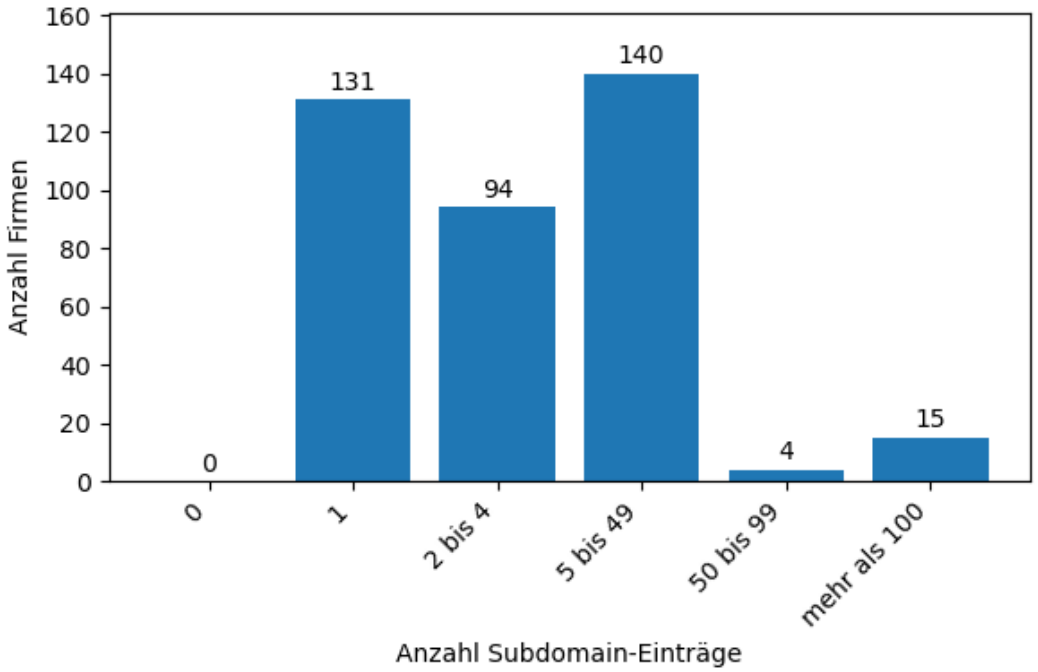

ABBILDUNG A.6: Subdomain-Records - Verteilung pro Firma

A.4.7 TXT-Records - Verteilung

Die Grafik zeigt die TXT-Verteilung. Bei 11 Firmen wurden 0 Einträge gefunden und bei 118 Firmen genau ein Eintrag. Bei 191 Firmen wurden bis zu vier Einträge gefunden. Bei 45 Firmen wurden bis zu zehn Einträge gefunden und bei 19 Firmen sogar mehr als zehn. Diese Verteilung ist nicht unüblich und technisch ist das kein Problem.

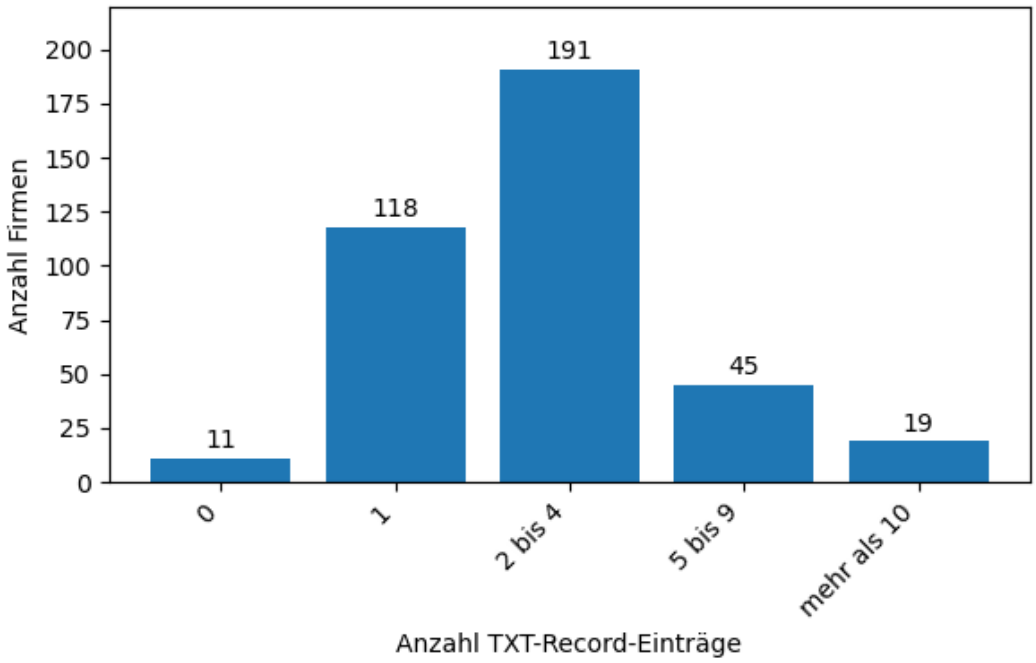

ABBILDUNG A.7: TXT-Records - Verteilung pro Firma

Literatur

[1] Bundesamt für Statistik. *Kleine und mittlere Unternehmen*. Zugriff = 03. Juni 2025. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/industrie-dienstleistungen/unternehmen-beschaeftigte/wirtschaftsstruktur-unternehmen/kmu.html.

[2] Alexandra Hüsler. *Digitale Kriminalität steigt weiter an*. Zugriff = 03. Juni 2025. URL: https://www.swisscybersecurity.net/news/2024-03-25/digitale-kriminalitaet-steigt-weiter-an?utm_source=chatgpt.com.

[3] Bundesamt für Statistik. *Digitale Kriminalität*. Zugriff = 03. Juni 2025. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/kriminalitaet-strafrecht/polizei/digitale-kriminalitaet.html?utm_source=chatgpt.com.

[4] Leonard Flach. *Hacker greifen überall und immer öfter an*. Zugriff = 03. Juni 2025. URL: https://www.srf.ch/news/schweiz/cyberattacken-in-der-schweiz-hacker-greifen-ueberall-und-immer-oefter-an.

[5] Frauenhofer Institut. *Was ist ein Digital Twin?* Zugriff = 03. Juni 2025. URL: https://www.ipk.fraunhofer.de/de/kompetenzen-und-loesungen/industrietrends/digital-twins.html#:~:text=Was%20ist%20ein%20Digital%20Twin,analysieren%2C%20simulieren%20und%20optimieren%20m%C3%B6chten..

[6] Rajiv Faleiro u. a. „Digital Twin for Cybersecurity: Towards Enhancing Cyber Resilience“. In: *International Conference on Broadband Communications, Networks, and Systems (BROADNETS 2021)*. Zugriff am 2. März 2025. Springer, 2022, S. 57–76. URL: https://link.springer.com/chapter/10.1007/978-3-030-93479-8_4.

[7] Moaiad Ahmad Khder. „Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application“. In: *International Journal of Advanced Soft Computing and Applications* (2021). Zugriff am 2. März 2025, S. 144–168. URL: https://www.researchgate.net/publication/357401723_Web_Scraping_or_Web_Crawling_State_of_Art_Techniques_Approaches_and_Application.

[8] Hack The Box Academy. *Information Gathering - Web Edition, Automate Recon*. Zugriff am 2. März 2025. URL: https://academy.hackthebox.com/module/144/section/3081.

[9] Hack The Box Academy. *Vulnerability Assessment*. Zugriff am 2. März 2025. URL: https://academy.hackthebox.com/module/108/section/1230.

[10] Universität Zürich. *Cyrenzh*. Zugriff am 12. Mai 2025. 2024. URL: https://cyrenzh.ch/de/project-de/.

[11] Prof. Dr. Bernhard Tellenbach. *PENETRATION TESTING (PART II)*. Vorlesungsfolie von Modul SWS2 Frühlingssemester 2025.

[12] Flutter. Zugriff = 04. Juni 2025. URL: https://flutter.dev/.
[13] Zugriff = 04. Juni 2025. URL: https://www.docker.com/.
[14] Ian Buchanan. Zugriff = 04. Juni 2025. URL: https://www.atlassian.com/de/microservices/cloud-computing/containers-vs-vms.
[15] Hack The Box Academy. *Penetrationstester Role Path*. Zugriff am 2. März 2025. URL: https://academy.hackthebox.com/path/preview/penetration-tester.
[16] Hack The Box Academy. *Information Gathering - Web Edition: DNS*. Zugriff am 28. Mai 2025. 2025. URL: https://academy.hackthebox.com/module/144/section/3074.
[17] Hack The Box Academy. *Information Gathering - Web Edition: Digging DNS*. Zugriff am 28. Mai 2025. 2025. URL: https://academy.hackthebox.com/module/144/%20section/1251.
[18] DataScientest. *Nmap einfach erklärt: Funktionen, Anwendungsbereiche und erste Schritte*. Zugriff am 28. Mai 2025. 2025. URL: https://datascientest.com/de/nmap-was-ist-das.
[19] Security Insider. *Was ist Metasploit?* Zugriff am 28. Mai 2025. 2025. URL: https://www.security-insider.de/was-ist-metasploit-a-688417/.
[20] Denys Kontorskyy. *DNS MX Records Explained*. Zugriff am 28. Mai 2025. 2023. URL: https://mailtrap.io/blog/dns-mx-records/.
[21] Heimdal Security. *Understanding DNS MX Records and Their Role in Email Security*. Zugriff am 28. Mai 2025. 2024. URL: https://heimdalsecurity.com/blog/understanding-dns-mx-records-and-their-role-in-email-security/.
[22] Hack The Box Academy. *Information Gathering - Web Edition: DNS Zone Transfers*. Zugriff am 28. Mai 2025. 2025. URL: https://academy.hackthebox.com/module/144/section/1255.
[23] James William Steven Parker. *What is reverse DNS lookup in the context of footprinting?* Zugriff am 28. Mai 2025. 2025. URL: https://www.cyberly.org/en/what-is-reverse-dns-lookup-in-the-context-of-footprinting/.
[24] Cloudflare. *What is a DNS SOA record?* Zugriff am 28. Mai 2025. 2025. URL: https://www.cloudflare.com/learning/dns/dns-records/dns-soa-record/.
[25] Rob Watts Shweta. *What Is A Subdomain? Everything You Need To Know*. Zugriff: 28. April 2025. Mai 2024. URL: https://www.forbes.com/advisor/business/what-is-a-subdomain/.
[26] Hack The Box Academy. *Information Gathering - Web Edition, Subdomains*. Zugriff am 28. April 2025. URL: https://academy.hackthebox.com/module/144/section/1252.
[27] Ahmed Aboul-Ela. *Sublist3r - Subdomain Enumeration Tool*. https://github.com/aboul3la/Sublist3r. Open Source Python-Tool zur Subdomain-Auflistung für Penetrationstests und Sicherheitsanalysen. 2020.
[28] *crt.sh - Certificate Transparency Search*. Zugriff am 28. Mai 2025. URL: https://crt.sh.
[29] Herbert Wieler. *NSA warnt vor Wildcard-Zertifikaten und Alpaca-Angriffen*. Zugriff am 28. Mai 2025. 2021. URL: https://www.infopoint-security.de/nsa-warnt-vor-wildcard-zertifikaten-und-alpaca-angriffen/a29067/.
[30] MDN Contributors. *What is a URL?* Zuletzt aktualisiert am 13. Mai 2025, Zugriff am 28. Mai 2025. 2025. URL: https://developer.mozilla.org/en-

US/docs/Learn_web_development/Howto/Web_mechanics/What_is_a_URL#summary.

[31] Artem Galan. *Angriffe über HTTP und wie man sich davor schützt*. Zugriff am 28. Mai 2025. 2024. URL: https://nine.ch/de/attacks-via-http-and-how-to-protect-yourself-against-them/.

[32] *Web Archives - Information Gathering - Web Edition*. Zugriff am 28. Mai 2025. 2025. URL: https://academy.hackthebox.com/module/144/section/1259.

[33] Internet Archive. *Wayback Machine*. Zugriff am 28. Mai 2025. 2025. URL: https://web.archive.org/.

[34] Stefan Pejčić. *List of TLD from IANA sorted by alphabetical order*. Zugriff am 28. Mai 2025. 2019. URL: https://gist.github.com/stefanpejcic/db876f13a28ec4021c98aa458541d68b.

[35] Hack The Box Academy. *Information Gathering - Web Edition: WHOIS*. Zugriff am 28. Mai 2025. 2025. URL: https://academy.hackthebox.com/module/144/section/3073.

[36] DeHashed. *DeHashed - Deep Web Search Engine for Leaked Data*. Zugriff am 28. Mai 2025. 2025. URL: https://dehashed.com/.

[37] Katherine Haan. „America's Password Habits: 46% Report Having their Password Stolen Over the Last Year". In: *Forbes Advisor* (2024). Zugriff am 28. Mai 2025. URL: https://www.forbes.com/advisor/business/americas-password-habits/.

[38] Faddom. *How Common Platform Enumeration Puts You in Control of IT Security*. Zugriff am 28. Mai 2025. 2024. URL: https://faddom.com/common-platform-enumeration/.

[39] Red Hat. *What is a CVE?* Zugriff am 28. Mai 2025. 2024. URL: https://www.redhat.com/en/topics/security/what-is-cve.

[40] Shodan. *Search Engine for the Internet of Things*. Zugriff am 28. Mai 2025. 2025. URL: https://www.shodan.io/.

[41] Shodan. *What is Shodan?* Zugriff am 28. Mai 2025. 2025. URL: https://help.shodan.io/the-basics/what-is-shodan.

[42] PostgreSQL Global Development Group. *PostgreSQL: The World's Most Advanced Open Source Relational Database*. Zugriff am 28. Mai 2025. 2025. URL: https://www.postgresql.org/.

[43] KMU Winterthur. Zugriff = 13. Main 2025. URL: https://www.kmu-win.ch/unternehmen-fur-sie/.

[44] Stack exchange. Zugriff = 22. Mai 2025. URL: https://superuser.com/questions/1018514/why-does-www-example-com-differ-from-example-com.

[45] Switch. *DNS-Abfragen: IPv4 im Vergleich zu IPv6*. Zugriff: 02. Juni 2025. URL: https://www.nic.ch/de/statistics/dns/ipv4-6/.

[46] Kinsta. *Was ist ein Nameserver? Warum sind Nameserver wichtig?* Zugriff = 02. Juni 2025. URL: https://kinsta.com/de/wissensdatenbank/was-ist-ein-nameserver/.

[47] cloudflare. *Was ist ein DNS-SOA-Eintrag?* Zugriff = 02. Juni 2025. URL: https://www.cloudflare.com/de-de/learning/dns/dns-records/dns-soa-record/#:~:text=Alle%20DNS%2DZonen%20ben%C3%B6tigen%20einen,sind%20auch%20f%C3%BCr%20Zonen%C3%BCbertragungen%20wichtig..

[48] cyon. Zugriff = 02. Juni 2025. URL: https://www.cyon.ch/.

[49] DeepL SE. *DeepL Write: AI Writing Assistant*. Zugriff: 14. April 2025. 2025. URL: https://www.deepl.com/en/write.

[50] DeepL Translate. Zugriff: 02. Juni 2025. URL: https://www.deepl.com/de/translator.

[51] OpenAI. *ChatGPT (GPT-4): Sprachmodell*. Zugriff: 14. April 2025. 2025. URL: https://chat.openai.com/.

[52] cloudflare. *Was ist ein DNS-PTR-Eintrag?* Zugriff = 02. Juni 2025. URL: https://www.cloudflare.com/de-de/learning/dns/dns-records/dns-ptr-record/.