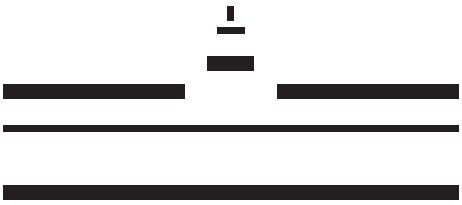


Westfälische
Wilhelms-Universität
Münster

Vergleich von klassischen und Cloud-basierten Netzwerk-Teleskopen anhand von UDP- und ICMP-Paketen

Bachelorarbeit

vorgelegt von:
Tobias Jutzi

Matrikelnummer: 525752

Studiengang: Informatik

Thema gestellt von:
Prof. Dr. Ralph Holz

Arbeit betreut durch:
Nils Kempen

Münster, 31. März 2025

Abstract

In dieser Arbeit werden die UDP-Pakete eines Cloud-Teleskops untersucht und gefundene Phänomene mit den Daten des CAIDA-Teleskops überprüft. Zusätzlich werden Verfahren vorgestellt, um für die Daten eines Cloud-Teleskops die zur Erkennung von UDP-Scans notwendigen Parameter eines minimalen Anfrage-Grenzwerts und eines zeitlichen Intervalls zu ermitteln. Drei Analysen visualisierten dabei die Verläufe der gefundenen Scans, der neu entdeckten Scanner und der durchschnittlichen Scandauer, sodass durch eine Sensitivitätsanalyse ein entsprechender Anfrage-Grenzwert bestimmt werden konnte. Durch die normalisierten Funktionen der zuvor beschriebenen Verläufe wird der Intervallwert ermittelt. Anhand der Daten des Cloud-Teleskops wird gezeigt, dass einzelne RIR-Namen für besonders aggressive Scan-Aktivitäten verantwortlich sind. Zudem wird gezeigt, dass sich die Scanstrategien der einzelnen Länder zwischen TCP und UDP unterscheiden. Bei der detaillierten Analyse einzelner Scans wird erkannt, dass einige Scanstrategien gezielt darauf ausgerichtet sind, Sicherheitsmechanismen zu umgehen. Die Untersuchung der gefundenen DoS/DDoS-Angriffe zeigt, dass der Ursprung der für einen Angriff verwendeten src-IPs keinen geografischen Zusammenhang zum angegriffenen Ziel aufweist. Zudem wird gezeigt, dass aggressive Scans und hierbei vor allem Port-Scans die IP-Adressen von CAIDA meiden. Abschließend wird der begründete Verdacht untermauert, dass Betreiber aus dem Iran gezielt Cloud-Maschinen mieten, um damit Address-Scans der jeweiligen Cloud-Infrastruktur durchzuführen.

Inhaltsverzeichnis

1 Einleitung **1**

2 Themenbezogende Arbeiten **3**

3 Hintergrund **5**
3.1 UDP . . . 5
3.1.1 DNS . . . 6
3.1.2 NTP . . . 7
3.2 ICMP . . . 7
3.3 Klassisches-Teleskop . . . 8
3.4 Cloud-Teleskop . . . 8
3.5 DoS/DDoS-Angriffe . . . 9
3.6 Scanning . . . 11

4 Methodik **13**
4.1 Zeek . . . 13
4.2 Beschreibung des verwendeten Cloud-Teleskops . . . 13
4.3 CAIDA-Teleskop . . . 14
4.4 Scan-Identifizierung . . . 15
4.5 DoS/DDoS-Identifizierung . . . 17
4.6 Geo-Datenbank . . . 18

5 Ergebnisse **21**
5.1 Cloud-Teleskop Allgemein . . . 21
5.2 Scanning . . . 22
5.2.1 Scan-Klassifizierung anhand des zeitlichen Intervalls und der Anzahl registrierter Verbindungen . . . 22
5.2.2 Zeitliche Betrachtung des Scan-Verhaltens . . . 27
5.2.3 Geografische Betrachtung des Scan-Verhaltens . . . 29
5.2.4 Genauere Betrachtung des Scan-Ursprungs . . . 32
5.3 DoS/DDoS-Angriffe . . . 35
5.4 Vergleich CAIDA-Teleskop . . . 39
5.4.1 Vergleich Scanning . . . 39
5.4.2 Vergleich DoS/DDoS-Angriffe . . . 41

6 Diskussion **45**
6.1 Herleitung der Parameter zur Identifizierung von Scans . . . 45
6.2 Bewertung der im Cloud-Teleskop gefunden Phänomene . . . 47
6.2.1 Bewertung Scanning . . . 47
6.2.2 Bewertung DoS/DDoS-Angriffe . . . 50
6.3 Bewertung der im CAIDA-Teleskop wiedergefundenen Phänomene 52

7 Zusammenfassung **55**

1 Einleitung

Die Nutzung von Cloud-Diensten sowohl für den privaten, als auch den geschäftlichen Bereich ist in den letzten Jahren stark gestiegen[1, 2]. Diese Verlagerung von Anwendungen und Daten in die Cloud bleibt auch von Kriminellen nicht unbeobachtet. So hat in den letzten Jahren die Anzahl an Angriffen gegen entsprechende Cloud-Umgebungen stark zugenommen[3]. Ein zentrales Element zur Untersuchung und Analyse des Netzwerkverkehrs sind Internet-Teleskope. So wurden seither klassische Internet-Teleskope dazu verwendet, kriminelle Aktivitäten aufzudecken und entsprechende Angriffsmuster zu verstehen[4]. Diese klassischen Teleskope greifen zum Sammeln der Daten auf ungenutzte IP-Adressen zurück[5]. Durch die oben beschriebenen Veränderung stellt sich allerdings die Frage, inwiefern diese Vorgehensweise noch zeitgemäß ist. Ein alternativer Ansatz, um Internet-Traffic aufzuzeichnen, ist die Verwendung der in den letzten Jahren aufgekommenen Cloud-Teleskope. Da diese Teleskope für das Sammeln der Daten die IP-Adressen der Cloud-Anbieter verwenden, ist so ein tieferer Einblick in den Datenverkehr dieser Cloud-Umgebungen möglich[1].

Um die Unterschiede dieser beiden Ansätze genauer zu untersuchen, wird in dieser Arbeit eine vergleichende Analyse beider Teleskop-Arten durchgeführt. Für diesen Vergleich wird auf je einen Datensatz der beiden Teleskope zurückgegriffen. Der Fokus liegt hierbei ausschließlich auf den in den Daten gefundenen UDP- und ICMP-Paketen. In diesem Zusammenhang werden die folgenden Forschungsfragen untersucht:

1. Durch welche Verfahren lassen sich für ein Cloud-basiertes Netzwerk-Teleskop die, zur Erkennung von Scans notwendigen, Parameter eines zeitlichen Intervalls sowie eines minimalen Grenzwerts an Anfragen ermitteln?

2. Welche Phänomene lassen sich in den UDP-Paketen eines Cloud-basierten Netzwerk-Teleskops identifizieren?

3. Inwiefern lassen sich die im Cloud-Teleskop gefundenen Phänomene in den Daten eines klassischen Teleskops wiederfinden?

Zur Beantwortung dieser Fragen werden empirische Analysen herangezogen. Der Fokus liegt hierbei auf den Themen Scanning und DoS/DDoS-Angriffe.

2 Themenbezogende Arbeiten

Ein Überblick themenbezogener Werke sowie die Einordnung dieser Arbeit in den aktuellen Forschungsstand wird im Folgenden Abschnitt dargestellt.

Moore *et al.*[4] gelten hierbei als einer der ersten, die umfangreiche Analysen zum Thema *backscatter-Traffic* durchführen. Durch den Einsatz klassischer Internet-Teleskope identifizieren sie internetweite *denial-of-service* Angriffe und beschreiben die hierzu notwendigen Analysemethoden. Zudem erklären sie, ähnlich wie Moore *et al.* in ihrer im Jahre 2004 veröffentlichten Arbeit[5], die Funktionsweise sowie den grundlegenden Aufbau eines klassischen Internet-Teleskops.

Die im Jahre 2017 von Blenn *et al.*[6] veröffentlichte Arbeit bietet einen weiteren sehr umfangreichen Überblick, über die im Internet stattfindenden DoS/DDoS-Attacken. Hierbei legen sie einen großen Fokus auf die Analyse der Intensität und Dauer identifizierter Angriffe. Erstmalig wenden sie hierzu ein Verfahren an, um mittels ICMP-Paketen die Dauer dieser Attacken zu bestimmen. Weiter finden sie heraus, dass die aus den Medien bekannten "mega Attacken"(100 Gbps und größer) nur einen kleinen Teil, der tatsächlich im Internet stattfindenden Angriffe ausmachen.

Speziell mit dem Thema *Amplification-Angriffe* befasst sich die Arbeit von Rossow[7]. In dieser beleuchtet er verschiedene Arten verstärkter DoS/DDoS-Angriffe und erklärt im Detail, welche Schwachstellen der jeweiligen Protokolle hierbei ausgenutzt werden.

Das Thema *Scanning* wird von Durumeric *et al.* in ihrer im Jahr 2014 erschienenen Arbeit[8] aufgearbeitet. Mittels eines klassischen Teleskops analysieren sie das Scan-Verhalten und beschreiben Veränderungen zu früheren Arbeiten. So finden sie heraus, dass sehr groß angelegte Scans nur von wenigen Ländern aus durchgeführt werden und diese Scans andere Ports anvisieren als kleinere. Zudem identifizieren sie die zur Durchführung von Scans eingesetzte Software und finden heraus, dass Tools wie ZMap und Masscan häufig für diese internetweiten Scans verwendet werden. Die Grundlagen sowie die verschiedenen Arten und Vorgehensweisen von Scans werden in der von Bhuyan veröffentlichten Arbeit[9] genauer betrachtet.

Der Frage, wie relevant/repräsentativ klassische Teleskope im Vergleich zu Cloud-Teleskopen in der heutigen Zeit noch sind, gehen Pauley *et al.* in ihrer im Jahr 2023 veröffentlichten Arbeit[1] nach. Hierzu vergleichen sie, ähnlich wie in dieser Arbeit, Daten eines klassischen Teleskops mit denen eines Cloud-Teleskops. Pauley und seinem Forschungsteam gelingt es hierfür ein Cloud-Teleskop zu entwickeln, welches in einer extrem kostengünstigen Art und Weise

eine beachtliche Datenmenge sammelt. Anhand dieser Daten zeigen sie, dass der Traffic ihres Teleskops von mehr einzigartigen IP-Adressen stammt (73% mehr unterschiedliche src-IPs) und dieser Traffic zudem eine größere Variabilität aufweist (95% höher). Weiter zeigen sie, dass die IP-Adressen des Cloud-Teleskops deutlich häufiger von Scannern verwendet werden. Zudem finden sie heraus, dass die geografische Lage des Scan-Ursprungs keinen signifikaten Zusammenhang mit der Region des Ziels aufweist.

Was, in der Arbeit von Pauley *et al.*, sowie in vielen anderen Studien nur oberflächlich oder gar nicht betrachtet wird, ist der UDP-Datenverkehr. Gerade die vermehrte Nutzung von *QUIC* in den letzten Jahren führt allerdings zu einer immer größeren Relevanz dieses Protokolls. Genau in diese Forschungslücke soll sich diese Arbeit einreihen und entsprechende Erkenntnisse liefern.

3 Hintergrund

Da in dieser Arbeit ausschließlich UDP- und ICMP-Traffic betrachtet wird, beginnt dieses Kapitel mit einem kurzen Überblick über beide Protokolle. Anschließend werden die Funktionsweisen und Unterschiede eines klassischen und Cloud-basierten Internet-Teleskops näher betrachtet. Abschließend erfolgt eine genaue Auseinandersetzung mit den Themen Scanning und DoS/DDoS-Angriffe, wobei vor allem auf die verschiedenen Scan- und Angriffsarten eingegangen wird.

3.1 UDP

UDP bildet neben TCP ein wichtiges Protokoll der Transportschicht des *ISO-OSI-Schichtenmodells*[10, 11]. Der Hauptunterschied dieser beiden Protokolle besteht darin, dass UDP verbindungslos ist. Dies bedeutet, dass vor der Datenübertragung keine stabile Verbindung zwischen den beiden Kommunikationspartner aufgebaut wird. Der Wegfall des vorgeschalteten Handshakes führt so zu einer schnelleren Übertragung der Daten, was UDP vor allem bei kleinen Datenmengen vorteilhaft macht. Dies bedeutet allerdings auch, dass die in der Anwendungsschicht laufenden Dienste dafür verantwortlich sind, die Konsistenz und Fehlerfreiheit der erhaltenen Pakete sicherzustellen und entsprechende Pakete gegenfalls erneut anzufordern. Für eine schnelle Datenübertragung spricht zudem das sehr simple aufgebaute UDP-Datagram. So enthält dies weder Informationen zur Flusskontrolle noch zur Staukontrolle. Lediglich eine einfache Prüfsumme ist enthalten. Aufgrund der oben genannten Eigenschaften wird UDP hauptsächlich in den Anwendungsfällen verwendet, in welchen eine schnelle Paketübertragung wichtiger ist als ein möglicher Paketverlust. Dies sind vor allem Echtzeitanwendungen wie Sprach- oder Videoanrufe[10].

Eine kurze Beschreibung einiger Protokolle der Anwendungsschicht, welche UDP als Transportprotokoll verwenden, wird im Rahmen der folgenden Unterkapitel gegeben. Hierbei wird vor allem auf einzelne Begriffe eingegangen, die für das Verständnis dieser Arbeit relevant sind, ohne die jeweiligen Protokolle in Gänze zu erklären.

3.1.1 DNS

Das *Domain Name System* wird verwendet, um zu einem Domainnamen die entsprechende IP aufzulösen. Hierzu werden *DNS-Server* verwendet. Diese sind hierarchisch strukturiert und haben zu den Domainnamen ihrer Hierarchieebene entsprechende *Resource-Record (RR)* Einträge gespeichert. Die zu einer Domain zugehörige IP wird hierbei als *A-Record* (IPv4) oder *AAAA-Record* (IPv6) gespeichert. Zusätzlich existiert eine Reihe weiterer Record-Typen, auf welche an dieser Stelle allerdings nicht weiter eingegangen wird. Die Kommunikation mit einem DNS-Server wird von einem *DNS-Resolver* durchgeführt, welcher DNS-Anfragen von Endgeräten entgegennimmt[10, 12]. Ein solcher (lokaler) DNS-Resolver ist in der Regel einem bestimmten IP-Adressbereich zugeordnet und akzeptiert ausschließlich Anfragen von Endgeräten dieses Bereiches. Als *offener DNS-Resolver* wird hingegen ein Resolver bezeichnet, welcher nicht über einen solchen Sicherheitsmechanismus verfügt und DNS-Anfragen von beliebigen IP-Adressen entgegennimmt[13]. Neben der zuvor beschriebenen Anfrage ist in dem DNS-Protokoll auch eine sogenannter *ANY-Request* definiert. Die Antwort einer solchen Anfrage enthält alle zu dem Domainnamen verfügbaren RR-Einträge[14]. Ein weiterer für das Verständnis der Arbeit relevanter Begriff ist das *DNSSEC*.

Die *Domain Name System Security Extensions* sind Erweiterungen, welche die Authentizität und Integrität in der DNS-Kommunikation gewährleisten. Hierzu wird mittels asymmetrischer Kryptografie ein *Public-Key-Signaturverfahren* angewandt. Jeder DNS-Server besitzt hierbei einen einzigartigen und geheimen Private-Key und einen öffentlichen Public-Key. Durch den Privat-Key und dem entsprechenden Hash-Wert eines DNS-Eintrages, wird für jeden dieser Einträge eine digitale Signatur erzeugt. Jede dieser Signaturen wird in einem neuen *RRSIG-Resource-Record* gespeichert, welches bei DNS-Anfrage als Antwort mitgegeben wird. Ein Resolver kann mittels des öffentlichen Schlüssels die digitale Signatur entschlüsseln und mit dem Dateninhalt vergleichen, um so die Echtheit zu garantieren. Das RRSIG-Record enthält typischerweise eine 1024 Bit lange Signatur und führt so zu erheblich größeren Antworten[15, 7].

Ein Protokoll, welches an späterer Stelle noch einmal aufgegriffen wird, ist das *Multicast-DNS (mDNS)* Protokoll. Dieses Protokoll wurde speziell für Geräte in lokalen Netzwerken entwickelt, welche nicht über einen entsprechenden DNS-Resolver verfügen. Ziel ist es, die Namensauflösung dieser Geräte untereinander per *Multicast-Anfragen* zu ermöglichen. So sendet ein Gerät eine mDNS-Anfrage an eine Multicast-IP (224.0.0.251 für IPv4), welche alle Geräte erhalten. Das zur Namensauflösung angefragte Gerät sendet nun seine IP per Multicast zurück. Wichtig zu beachten ist, dass für solche Multicast-Anfragen kein Authentifizierungsmechanismus oder ähnliches verwendet wird. Ähnlich wie oben enthält auch dieses Protokoll einen *ANY*-Abfragetyp. Hier führt eine solche ANY-Abfrage dazu, dass alle Geräte des lokalen Netzes mit all ihren gespeicherten mDNS-Namensauflösungen antworten[16, 17, 18].

3.1.2 NTP

Das *Network Time Protocol* wird zur Synchronisation von Uhrzeiten in Netzwerken verwendet. Hierzu teilt der Client dem *NTP-Server* mittels eines NTP-Request den Zeitstempel mit, bei welchem er die Anfrage versendet (*Originate-Timestamp*). Der Server antwortet dem Client mittels NTP-Response und übermittelt neben dem erhaltenen Zeitstempel die folgenden Werte:
- *Receive-Timestamp*: Empfangszeitpunkt der Anfrage
- *Transmit-Timestamp*: Zeitpunkt des Verschickens der Antwort
Der Zeitpunkt, an welchem der Client die Antwort erhält, ist der *Destination-Timestamp*. Durch das Einsetzen der Werte in die folgende Formel, ergibt sich der *Clock-Offset*, um welchen die Uhrzeit des Clients aktualisiert wird[19, 20]:

$$\text{Clock Offset} = \frac{(Receive - Originate) + (Transmit - Destination)}{2}$$

Ein weiterer Anfragetyp, welcher standardmäßig in älteren NTP-Versionen enthalten ist, ist der sogenannte *monlist-request*. Ursprünglich für die Serveradministration und für Statusabfragen entwickelt, gibt die monlist-Anfrage eine Liste mit den letzten 600 IP-Adressen zurück, welche zuvor den Server angefragt haben[7, 21]. Da dies, wie in Abschnitt 3.5 zu sehen, zu einigen Sicherheitsrisiken führt, ist dieser Anfragetyp ab der NTP-Version 4.2.7 standardmäßig deaktiviert[21].

3.2 ICMP

Das *internet-control-message-protocol* gehört im Gegensatz zu TCP/UDP nicht zur Transportschicht, sondern ist über dem IP-Protokoll in der Netzwerkschicht angeordnet. ICMP-Nachrichten werden sowohl von Endgeräten als auch von Routern verwendet und als IP-Dataframe verschickt. Durch einen Typ, einen zugehörigen Code und eine Beschreibung können so Zustandsinformationen der Netzwerkschicht ausgetauscht werden. So kann ein Router mittels der Beschreibung *source quench* (Type 4 und Code 0) dem Sender eine volle Eingangswarteschlange, also eine Verstopfung, mitteilen. Ähnlich kann ein Server auf einen *echo-request* Ping (Type 8 und Code 0) mit einem *echo-reply* (Type 0 und Code 0) antworten. Dieser echo reply ist notwendig, da eine empfangene ICMP-Nachricht im Allgemeinen keine Bestätigung verschickt wird[10]. Ein weiteres ICMP-Szenario, welches später in dieser Arbeit noch einmal aufgegriffen wird, ist das Folgende:

Empfängt ein Server eine TCP-SYN-Anfrage, so würde er bei einem offenen Port mit einem SYN+ACK antworten. Ist der Server allerdings überlastet (da er aktuell zu viele Anfragen erhält), so antwortet er in einigen Implementierungen mit einem ICMP *destination unreachable* (Type 3 und Code 1)[6].

3.3 Klassisches-Teleskop

Ein klassisches Internet-Teleskop (im weiteren Verlauf auch als Darknet-Teleskop bezeichnet) ist ein System, welches es ermöglicht, eingehenden Internetverkehr bestimmter IP-Adressen aufzuzeichnen. Bei den von einem Darknet-Teleskop verwendeten IP-Adressen, handelt es sich um unbenutzte Adressen, welche oft aus einem zusammenhängenden IP-Adressraum stammen. Unbenutzte IP-Adressen bedeutet in diesem Zusammenhang, dass diese von keinen legitimen Endgeräten verwendet werden und somit auch niemand einen sinnvollen Nutzen darin hat, Anfragen an diese zu senden. Zur Aufzeichnung des Netzwerkverkehrs wird ein *Border-Router* so konfiguriert, dass sämtlicher eingehender Traffic, welche in dem IP-Adressraum des Teleskops liegt, an dessen Infrastruktur weitergeleitet wird. Um diese Umleitung durchzuführen, benötigten die Besitzer des Teleskops entweder selber die Rechte an dem IP-Adressraum oder sie kooperieren mit entsprechenden Dritten. Zwei wichtige Bestandteile der Teleskop-Infrastruktur sind eine leistungsfähige (CPU starke) Recheneinheit sowie ein großes Speichersystem. Die Recheneinheit wird verwendet, um den eingehenden Traffic in Echtzeit zu verarbeiten. Hierzu zählt das vorfiltern und aufbereiten, sowie das Erzeugen geeigneter Datentypen/Pakettypen. Die erzeugten Daten (oft im Pcap-Format) werden anschließend in einem Datenbanksystem gespeichert. Wichtig ist hier eine organisierte Speicherung, sowie ein schneller (leistungsstarker) Zugriff auf die gespeicherten Daten[22]. Einige Teleskop-Infrastrukturen ermöglichen zudem die Echtzeitanalyse des eingehenden Netzwerkverkehrs. Hierdurch ist ein live-Monitoring der erfassten Daten möglich. Um dies Umzusetzen ist ein spezieller *Real-Time-Packet-Stream* notwendig, welche den Zugriff in Echtzeit ermöglicht[23]. Die wichtigste Kenngröße eines Internet-Teleskops ist der zugrundeliegende IP-Adressbereich. Gesammelte Netzwerkdaten, welche auf vielen unterschiedlichen IP-Adressen basieren, führen zu aussagekräftigeren Ergebnissen. Zudem können so Internetweite Phänomene identifiziert werden (z.B. internetweites Scanning). Die Größe wird hierbei üblicherweise in der *CIDR*-Schreibweise angegeben, wobei \X bedeutet, dass 32 - X Bit für den *host part* verwendet werden, also 2^{32-X} verschiedene IP-Adressen. Die Abdeckung des Teleskops ($\hat{T}$) von dem gesamten 2^{32} großen IPv4 Adressbereich errechnet sich somit wie folgt[22]:

$$\hat{T} = \frac{1}{2^x}$$

3.4 Cloud-Teleskop

Der Aufbau eines Cloud-Teleskops, welcher im Folgenden beschrieben wird, basiert hauptsächlich auf der Implementierung aus[1]. Dies ist damit zu begründen, dass das Cloud-Teleskop, welches für diese Arbeit verwendet wurde, diesem Aufbau sehr ähnelt (siehe Abschnitt 4.2).

So wie ein Darknet-Teleskop, wird auch ein Cloud-Teleskop dazu verwendet, Internet-Traffic aufzuzeichnen. Der größte Unterschied der beiden Teleskope besteht darin, auf welche IP-Adressen zum Sammeln der Daten zurückgegriffen wird. Ein Cloud-Teleskop verwendet IP-Adressen aus dem IP-Adressbereich eines oder mehrerer Cloud-Anbieter. Um den Internet-Traffic dieser IP-Adressen aufzuzeichnen, werden temporär kleine Recheninstanzen des Cloud-Anbieters gemietet. Um die Mietkosten hierfür kleinzuhalten und dadurch, dass weder viel Rechenleistung noch Speicher benötigt wird, werden oft solche Instanzen verwendet, welche auf einen geteilten Speicher sowie Rechenleistung zugreifen. Wichtig ist, dass diese Instanzen eine feste IP-Adresse aus dem Adressbereich des Cloud-Anbieters besitzen. Der gesamte ankommende Traffic dieser IP-Adressen wird dann durch die jeweilige Instanz an einen dedizierten Service umgeleitet. Dieser ist dafür verantwortlich, mit den eingehenden Anfragen zu interagieren und diese zu speichern (Mehr zum Thema Interaktion mit eingehenden Anfragen in Abschnitt 4.2). Zur Speicherung werden die Daten in aggregierter Form in einer Datenbank abgelegt. Neben dieser Aggregation verfügen einige Cloud-Teleskope über Funktionalitäten, welche es ermöglichen, die aufgezeichneten Pcap-Daten in Echtzeit mit zusätzlichen Informationen zu erweitern. Dies kann beispielweise eine DNS-Namensauflösung sein.

Die optimale Mietdauer der einzelnen Recheninstanzen ergibt sich daraus, bei möglichst geringen Kosten ausreichend Traffic auf möglichst vielen verschiedenen IP-Adressen aufzuzeichnen. Cloud-Teleskope sind im Gegensatz zu klassischen Teleskopen verhältnismäßig einfach skalierbar. Dies ist dadurch zu begründen, dass die einzelnen Recheninstanzen nicht voneinander abhängen und diese automatisiert erstellt und gelöscht werden. Um eine breite (internetweite) Messung durchführen zu können, ist bei der Auswahl des Cloud-Anbieters darauf zu achten, dass dieser Rechenzentren in vielen Regionen/Kontinenten betreibt[1].

3.5 DoS/DDoS-Angriffe

Bei einer DoS/DDoS-Attacke versucht der Angreifer die legitime Nutzung eines Dienstes zu stören bzw. zu unterbrechen. Zur Erreichung dieses Ziels wird zwischen zwei Vorgehensweisen unterschieden:

1. Der Angreifer sendet eine erhebliche Anzahl an Anfragen an das Opfer, sodass die Rechenleistung nicht ausreicht, um alle diese Anfragen zu bearbeiten. Hierbei zielt der Angreifer darauf ab, dass für jede eingehende Anfrage ein neuer Prozess gestartet wird. So wird entweder versucht die CPU des Servers zu überlasten oder die maximal mögliche Anzahl an Verbindungen, welche der Server aufbauen kann, auszureizen[6].

2. Der Angreifer versucht nicht den Server selber zu attackieren, sondern zielt auf dessen Internetverbindung ab. Das Ziel besteht darin, durch das massenhafte Senden von Datenpaketen die Bandbreite der Verbindung zu überlasten und so den legitimen Datenverkehr zu blockieren[6].

Der Unterschied zwischen DoS (Denial of Service) und DDoS (Distributed Denial of Service) Attacken besteht darin, von wie vielen Geräten aus Anfragen an das Opfer gesendet werden. Im Gegensatz zu einem DoS-Angriff, bei welchem die Anfragen von einem einzelnen Gerät ausgehen, wird bei einem DDoS-Angriff auf ein ganzes Netzwerk an Geräten zurückgegriffen. Ein solches Netzwerk, bestehend aus vielen im Vorfeld infizierten Geräten, wird als *Botnet* bezeichnet. Diese Geräte sind so infiziert, dass sie ferngesteuert für eine solche Attacke eingesetzt werden können[24]. Zum Infizieren der Geräte werden Würmer wie *Mira* oder *Reaper* verwendet. Durch das Scannen und Ausnutzen von Sicherheitslücken sind diese Würmer in der Lage, sich selbst zu verbreiten. Dies führt zur Entstehung von entsprechend großen Botnets, welche in der Lage sind, enorme Datenlasten beim Opfer zu erzeugen[25, 6]. Da diese DDoS-Attacken von vielen verschiedenen Geräten aus durchgeführt werden, ist es für Sicherheitssysteme schwieriger, diese von legitimen Anfragen zu unterscheiden[24]. Um bei einem Angriff die eigene Identität zu verbergen, setzen Angreifer auf das sogenannte *IP-Spoofing*. Hierbei ersetzt der Angreifer die src-IP der Datenpakete durch eine (zufällig) generierte IP-Adresse. Dieses Verfahren wird häufig bei DoS-Attacken eingesetzt, um so die Erkennung für Sicherheitssysteme zu erschweren[6]. Wie unten bei der Beschreibung der verschiedenen Angriffsarten nachzulesen ist, existieren weitere Möglichkeiten, um mittels IP-Spoofing Angriffe durchzuführen.

Die bei einem Angriff verwendeten Strategien lassen sich wiederum in zwei Kategorien unterteilen. Dies sind zum einen *protokollbasierte-Angriffe*, welche darauf ausgelegt sind, Schwachstellen in einem Protokoll auszunutzen und *volumenbasierte-Angriffe*, welche sich allein durch eine enorme Datenlast auszeichnen[6]. Im Folgenden nun einige UDP-basierte Angriffsstrategien:

UDP-Flooding: UDP-Flooding ist ein Standardbeispiel für volumenbasierte Angriffe. Das Opfer wird mit einer so großen Menge an UDP-Paketen konfrontiert, sodass die Rechenleistung nicht ausreicht, um all diese Pakete zu bearbeiten. Für diese Art von Angriffen können alle Arten von UDP-Paketen verwendet werden[24].

NTP-Amplification: Diese Angriffsart gehört zu den protokollbasierten Angriffen, da hier eine Schwachstelle im NTP-Protokoll ausgenutzt wird. Wie in Unterabschnitt 3.1.2 beschrieben, enthält das NTP-Protokoll die monlist-Anfrage, welche als Antwort eine Liste der letzten 600 Clients zurückgibt. Die so erzeugte Antwort besteht aus bis zu 100 UDP-Datagrams, von denen jedes über einen Payload von ≈440 Bytes verfügt. Die so generierte Antwort ist um ein Vielfaches größer als die zugehörige Anfrage (≈8 Bytes). Durch IP-Spoofing wird die Source-IP der monlist-Anfrage mit der IP-Adresse des Opfers ersetzt und die Anfrage an einen regulären (offenen) NTP-Server ge-

stellt. Der NTP-Server dient als Verstärker und das Opfer enthält ausgehend von einer sehr kleinen Anfrage eine deutlich größere Antwort, welche so zur Überlastung führen soll.[7]

DNS-Amplification: Diese Angriffsart ähnelt der vorherigen und gehört ebenfalls zu den protokollbasierten-Angriffen. Als Verstärker dienen hier offene DNS-Resolver oder autoritative DNS-Server. Wie in Unterabschnitt 3.1.1 beschrieben, erzeugen diese auf ANY-Anfragen sehr große Antworten, da in der Antwort alle zu der Domain verfügbaren RR-Einträge enthalten sind. Die im selben Unterabschnitt beschriebenen DNSSEC-Signaturen bieten eine weitere Möglichkeit, um durch kleine Anfragen verhältnismäßig große Antworten zu generieren[7].

3.6 Scanning

Internet-Scanning in seiner grundlegenden Form beschreibt ein Verfahren, um herauszufinden, welche Dienste auf einem Host bzw. in einem Netzwerk aktiv sind. Spezialisierte Scans können zusätzlich weiterführende Informationen über den laufenden Dienst liefern. Hierbei ist das grundlegende Verfahren eines aktiven Scans immer dasselbe. Zu einem Port eines Hosts wird eine Anfrage gesendet und die darauf erhaltene Antwort analysiert. Dieses Verfahren, in wiederholter Form angewandt, lässt sich in drei Scan-Typen einteilen:

- *Port-Scan*: Ein Scan lässt sich diesem Typ zuordnen, wenn dieser darauf abzielt, bei einem einzelnen Host viele verschiedene Ports anzufragen.

- *Address-Scan*: Diese Art von Scan liegt vor, wenn bei vielen verschiedenen Hosts ein einziger vorher ausgewählter, Port angefragt wird.

- *Mixed-Scan*: Dieser auch als *Strobe-Scan* oder *Random-Scan* bezeichneter Scan-Typ beschreibt die Kombination der oberen beiden Typen. Auf mehreren Hosts wird eine Vielzahl von Ports angefragt.

Scans können sowohl auf TCP- als auch auf UDP-Ports durchgeführt werden, wobei TCP deutlich häufiger betroffen ist. Dies kann mit dem in Abschnitt 3.1 beschriebenen Unterschied der beiden Protokolle begründet werden. Da UDP im Gegensatz zu TCP verbindungslos ist, erhalten die Scan-Antworten weniger Informationen.

Durch das Aufbauen einer TCP-Verbindung, also dem Durchführen des 3-Wege-Handshakes, erhält der Sender so bei einem klassischen *SYN-Scan* in jedem Fall eine Antwort. Bei diesem Scan wird das *SYN-Flag* wie bei einem gewöhnlichen Verbindungsaufbau gesetzt. Ist der Port erreichbar, so enthält die Antwort das *ACK-Flag*. Bei einem geschlossenen Port wird hingegen das *RST-Flag* in der Antwort gesetzt[9]. Weitere in einem TCP-Paket enthaltende Informationen sind das *Receive Window* sowie die *Maximum Segment Size*[10].

Antworten von UDP-Scans enthalten im Gegensatz dazu deutlich weniger Informationen. So erhält der Sender durch eine nicht weiter spezifizierte UDP-Anfrage an einen offenen Port in der Regel keine Antwort. Bei einem geschlossenen Port wird ein ICMP *Port unreachable* (Type 3) Paket als Antwort gesendet[9]. Weiterreichende Informationen erhalten Scans durch spezifische UDP-Anfragen, welche von dem Dienst des jeweiligen Ports akzeptiert werden. Die so erhaltenden Informationen können von Angreifern genutzt werden, um verwundbare Ziele zu identifizieren und mögliche offene Ports zu ermitteln. So werden häufig die in Unterabschnitt 3.1.1 und Unterabschnitt 3.1.2 beschriebenen Anfragen an entsprechende NTP- / DNS-Ports versendet, um so offene Server zu identifizieren. Aus diesem Grund werden entsprechende Scans häufig den in Unterabschnitt 5.4.2 beschriebenen Angriffen vorgeschaltet[10, 8].

4 Methodik

In diesem Kapitel werden die Methoden und Verfahren erläutert, die zur Gewinnung der in Kapitel 5 dargestellten Ergebnisse verwendet wurden.

4.1 Zeek

Zur Umwandlung der von den Teleskopen aufgezeichneten Pcap-Dateien in menschenlesbares Format wurde in dieser Arbeit das Kommandozeilentool *Zeek* verwendet. Hierbei erstellt Zeek für jedes Protokoll der Anwendungsschicht eine Log-Datei, welche die zu diesem Protokoll gefundenen Verbindungsdaten enthält. Zudem werden weitere Log-Dateien für verdächtige (*notice.log*) oder merkwürdige (*weird.log*) Verbindungen erzeugt. Die zusätzlich erstellte *conn.log* enthält den gesamten Traffic und bündelt hierbei alle Paketdaten einer Verbindung in einem Eintrag[26]. Die in der späteren Analyse verwendeten Daten stammen hauptsächlich aus den Log-Dateien: conn.log, dns.log, ntp.log und notice.log. Zur Erkennung von Protokollen der Anwendungsschicht greift Zeek nicht nur auf Port basierten Heuristiken zurück, sondern verwendet auch inhaltsbasierte Signaturen, um den Payload anhand bekannter Protokollmuster zu überprüfen[27].

Da Zeek auf einer komponentenbasierten Architektur aufbaut, kann das Parsing problemlos durch zusätzliche Scripte erweitert werden. Für diese Arbeit wurden zu der standardmäßigen Konfiguration die folgenden von Zeek mitgelieferten Scripte eingebunden: auth-addl.zeek, weirds.zeek, speculative-service.zeek und unknown-protocols.zeek[28]. Zudem wurde zur Erkennung von UDP-Scans das benutzerdefinierte Script *scan_udp.bro*[29] in abgewandelter Form verwendet.

4.2 Beschreibung des verwendeten Cloud-Teleskops

Das für diese Arbeit verwendete Cloud-Teleskop stammt aus dem Forschungsbereich von Holz des Instituts für Informatik der Universität Münster[30]. Dieses Cloud-Teleskop wurde von dem Forschungsbereich für den Analysezeitraum von 20.01.2025 um 16:00 bis zum 03.02.2025 um 01:33 betrieben. Die gemieteten Cloud-Maschinen des Teleskops stammen hierbei von dem Cloud-Anbieter Digital-Ocean[31]. Das Teleskop wurde so skaliert, dass über den

gesamten Analysezeitraum zeitgleich 43 Cloud-Maschinen gemietet wurden. Die Betriebsdauer jeder Maschine betrug hierbei sechs Stunden. Nach Ablauf dieser Zeit wurde die entsprechende Maschine beendet und durch eine neue ersetzt. Somit ergibt sich die maximale Anzahl unterschiedlichen src-IPs des Teleskops wie folgt:

$$2.408 \text{ src-IPs} = \frac{43 \text{ Maschinen} \times 14 \text{ Tage} \times 24 \text{ Stunden}}{6 \text{ Stunden Betriebsdauer}}$$

Die verwendeten Cloud-Maschinen werden in den folgenden Ländern betrieben: USA, Singapur, Niederlande, Indien, Großbritannien, Deutschland, Kanada und Australien. Das Teleskop erzeugt für jede verwendete src-IP einen neuen Ordner mit den zugehörigen Pcap-Dateien. In jedem dieser Ordner ist zudem eine *descriptor.txt* Datei enthalten, welche Metadaten der jeweiligen Cloud-Maschine enthält. Dies sind Metadaten wie *hostname*, *creation*, *region* oder *deletion*. Dem Autor dieser Arbeit wurde der vollständige Zugriff auf die zuvor beschriebenen Ordner gegeben.

Wie in Abschnitt 3.4 bereits angesprochen, können Cloud-Teleskop entsprechend konfiguriert werden, um mit eingehenden Anfragen zu interagieren. Im Falle des verwendeten Cloud-Teleskops bedeutet dies, dass auf alle TCP-Anfragen mit einem TCP-RST geantwortet wird. UDP-Anfragen werden hingegen mit einem ICMP *Port unreachable* (Type 3) beantwortet. Ping-Anfragen (Type 8 und Code 0) werden mit dem vorgesehenen *echo-reply* (Type 0 und Code 0) beantwortet.

Weitere Informationen über die genaue Funktionsweise sowie den Aufbau des Teleskops liegen dem Autor nicht vor und können an dieser Stelle nicht näher erläutert werden.

4.3 CAIDA-Teleskop

Um die im Cloud-Teleskop identifizierten Phänomene mit Daten eines Darknet-Teleskops zu vergleichen, wurde für diese Arbeit das *UCSD-Network-Telescope* verwendet. Dieses Darknet-Teleskop sammelt den Internet-Traffic von einem /9 und einem /10 IP-Adressbereich und gehört somit zu den größten Internet-Teleskopen weltweit. Durch das Teleskop werden ungefähr 1/256 (≈12 Millionen) aller IPv4 Adressen überwacht[32, 23].

Das *CAIDA-Stardust* Projekt umfasste eine Sammlung von Tools und Infrastrukturen um auf die Daten des Teleskops zuzugreifen. So werden beispielweise die Daten als *Flow-Tupel* Format sowie als *Traffic-Traces* Format in einer Swift Datenbank bereitgestellt[23]. Die Traffic-Traces Daten sind nicht anonymisierte Pcap-Dateien, welche den aufgezeichneten Datenverkehr von jeweils einer Stunde beinhalten und somit ca. 100 GB groß sind. Diese Dateien werden 30 Tage lang auf der Swift-Datenbank bereitgestellt[33]. Bei dem Flow-Tupel Format werden die Daten hingegen in aggregierten Form gespeichert. So werden ähnliche Verbindungen in einem *Flow-Tupel-Record* zusammengefasst.

Dies reduziert den Speicherbedarf erheblich. Schlüsselwerte für diese Aggregation sind Quell-IP, Ziel-Subnetz, Ziel-Port sowie das Protokoll. Neben den Schlüsselwerten beinhalten die Flow-Tupel zahlreiche Metadaten der zusammengefassten Verbindungen[34].

Um eine möglichst hohe Vergleichbarkeit zu den Daten des Cloud-Teleskops herzustellen, wurden für diese Arbeit die Traffic-Traces Daten verwendet. Der Zugriff auf die Swift-Datenbank erfolgte mittels einer vom CAIDA-Stardust Projekt bereitgestellten VM. Zum Laden der Daten wurden einige Shell-Skripte sowie das *Libtrace-Toolset* verwendet. Dieses Toolset ermöglicht einen einfachen Zugriff auf den Stardust-Cloudspeicher. Da die verwendete VM über einen begrenzten Speicherplatz verfügt, wurde durch das Toolset zudem eine Vorfilterung der zu ladenden Daten vorgenommen. Diese Filterung erfolgte anhand einer Liste an IP-Adressen und dem Protokollfilter UDP[35, 36, 37].

Die grundlegende Funktionalität der Scripte besteht darin, die rohen Pcap-Dateien nacheinander zu laden und in jedem Iterationsschritt direkt mit Zeek zu verarbeiten. Die verwendeten Skripte basieren auf denen, welche von Kempen für eine Masterarbeit[38] entworfen wurden. Erweitert wurden diese Scripte mit der Funktionalität, durch ein als Parameter mitgegebenes Zeitfenster, vorzufiltern, welche Traces-Daten geladen werden sollen. Um den Speicherbedarf zu verringern, wurde zudem eine Filterung für die von Zeek erzeugten conn.log und notice.log entworfen. So werden in jedem Iterationsschritt die Dateien so verkleinert, dass nur im Vorfeld definierte Spalten erhalten bleiben. Das Ergebnis der Scripte sind konkatenierte Log-Dateien, welche aus den von Zeek in jedem Iterationsschritt erzeugten Teilstücken bestehen. Um vergleichbare Daten der beiden Teleskope zu erhalten, wurde der Parameter des Zeitfensters so gesetzt, dass nur CAIDA-Daten aus dem gleichen Analysezeitraum betrachtet werden, in welchem das Cloud-Teleskop betrieben wurde.

4.4 Scan-Identifizierung

Im Folgenden werden nun die Verfahren beschrieben, welche auf die Daten beider Teleskope zur Identifizierung von Scans angewandt wurden. Um mögliche Scans in einer Pcap-Datei zu erkennen, sind zwei Parameter von entscheidender Bedeutung. Dies sind zum einen ein zeitliches Intervall und zum anderen ein minimaler Grenzwert. Eine Quell-IP, welche Anfragen sendet, wird nur dann als Scan identifiziert, sobald die Anzahl gesendeter Anfragen den Grenzwert überschreitet. Zudem muss für jede dieser Anfragen die Bedingung erfüllt sein, dass zwei aufeinanderfolgende Anfragen zeitlich nicht weiter auseinanderliegen als das gesetzte Intervall. Der beschriebene Grenzwert wird im weiteren Verlauf dieser Arbeit als *Anfragen-Grenzwert* bezeichnet. Wie in Abschnitt 3.6 beschrieben, gelten weitere spezifische Einschränkungen für die Unterteilung in Port- und Address-Scan:

Bei einem Port-Scan muss für alle Anfragen des betrachteten Grenzwerts gelten, dass diese an eine einzige Ziel-IP gestellt werden und hierbei jede der Anfragen an einen unterschiedlichen Port gerichtet ist.
Um als Address-Scan identifiziert zu werden, muss hingegen gelten, dass alle Anfragen an unterschiedliche Ziel-IPs gerichtet sind und hierbei jeweils der gleiche Port angefragt wird.

Im weiteren Verlauf dieser Arbeit wird in diesem Zusammenhang zwischen den folgenden Begriffen unterschieden:

- *Scan-Anfrage*: Beschreibt eine einzelne Anfrage, welche zu einem Scan gehört.
- *Scan*: Beschreibt die Bündelung mehrerer Scan-Anfragen, welche zusammen die oben genannten Bedingungen erfüllen.
- *Scanner*: Wird durch eine eindeutige Quell-IP definiert. Alle Scans dieser Quell-IP gehören zu dem Scanner.

Zur Erkennung von möglichen Scans wurde für diese Arbeit in einem ersten Schritt das benutzerdefinierte Zeek-Script *scan_udp.bro*[29] verwendet. Dieses Script verfügt über die Möglichkeit, ein entsprechendes Intervall sowie einen Anfragen-Grenzwert als Parameter im Code zu setzen. Anhand dieser Parameter werden mögliche Scans in einer Pcap-Datei gefunden und in der notice.log ausgegeben. Da es sich bei diesem Script um ein benutzerdefiniertes Script handelt, welches von der Zeek-Community entworfen wurde, ist die Korrektheit des Codes nicht konkret geprüft. So konnten in der erzeugten notice.log einige Unstimmigkeiten erkannt werden. Zur weiteren Überprüfung des Scripts wurden mittels Python einige Test-Pcap-Dateien erstellt, welche ein konkretes Scan-Szenario simulieren[39]. Dadurch konnte festgestellt werden, dass die Bedingung des Anfragen-Grenzwertes korrekt im Script geprüft wird. Weiter konnte allerdings nachgewiesen werden, dass die Überprüfung des zeitlichen Intervalls nicht korrekt durchgeführt wird. Dies führt dazu, dass Scans als solche identifiziert werden, bei welchen zwei aufeinanderfolgende Anfragen weiter als das vorgegebene Intervall auseinanderliegen. Diese falsch-positive Erkennung führt somit dazu, dass in der erzeugten notice.log alle tatsächlichen Scans enthalten sind, aber auch solche, für welche die Bedingung des zeitlichen Intervalls nicht erfüllt ist. Ein natives Zeek-Script[40] existiert nur zur Identifizierung von TCP-Scans. Da dieses Script auf die Flags des TCP-Handshakes zurückgreift, würde eine Anpassung für UDP einen erheblichen Mehraufwand darstellen. Aus diesem Grund wurde zur Identifizierung von UDP-Scans ein Python-Algorithmus entworfen, welcher allerdings Informationen aus der zuvor genannten notice.log verwendet. Da, wie oben beschrieben, die Überprüfung des Anfragen-Grenzwertes korrekt in dem Script durchgeführt wird, verwendet der Algorithmus die Quell-IPs dieser notice.log als eine Art Vorfilterung. Dies trägt zu einer erheblich kürzeren Laufzeit bei. Neben den Quell-IPs der

notice.log verwendet der Algorithmus die Verbindungsdaten aus der in Abschnitt 4.1 beschriebenen conn.log.

Der grundlegende Programmablauf besteht darin, dass in einem ersten Schritt für jede dieser Quell-IPs alle zugehörigen Einträge der *conn.log* iterativ durchlaufen werden. Hierbei wird für alle zugehörigen Verbindungen die Bedingung des zeitlichen Intervalls überprüft. In einem weiteren Schritt wird anhand der oben beschriebenen Bedingungen die Unterteilung in Port- und Address-Scan durchgeführt. Der Algorithmus ist hierbei so aufgebaut, dass der Intervall-Wert als Parameter gesetzt werden kann.

Für die Daten des Cloud-Teleskops wurde zur Identifizierung von Port- sowie Address-Scans ein Anfragen-Grenzwert von 6 gewählt. Für Port-Scans wurde das zeitliche Intervall auf 7 Minuten und für Address-Scans auf 6 Minuten festgelegt. Eine ausführliche Betrachtung und Herleitung dieser Werte findet sich in Unterabschnitt 5.2.1 sowie Abschnitt 6.1. Die durch diese Parameter gefundenen Scans werden im Folgenden dieser Arbeit als *address-scan.log* und *port-scan.log* bezeichnet.

Um Address-Scans in den Daten des CAIDA-Teleskops zu identifizieren, wurde ein Anfragen-Grenzwert von 100 sowie ein zeitliches Intervall von 8 Minuten verwendet. Diese Werte stammen aus Studien, welche Daten eines ähnlichen Teleskops untersucht haben[8, 41]. Für Port-Scans wurde der Anfragen-Grenzwert von 25 bei gleichem zeitlichem Intervall festgelegt. Dieser Anfragen-Grenzwert stammt aus dem nativen Zeek-Script[40] zur Erkennung von Scans.

4.5 DoS/DDoS-Identifizierung

Im Folgenden nun eine Beschreibung der Vorgehensweise, wie in den Daten der beiden Teleskope mögliche DoS/DDoS-Angriffe identifiziert wurden. Ein fundamentaler Begriff in diesem Zusammenhang ist der sogenannte *Backscatter-Traffic*. Als Backscatter-Traffic werden die Antwortpakete bezeichnet, welche auf Anfragen von IP-Spoofing zurückzuführen sind[6]. Im Kontext von Internet-Teleskopen bedeutet dies, dass die gefälschte src-IP der Anfrage in dem Adressbereich des Teleskops liegt. Im Zusammenhang mit den in Abschnitt 3.5 beschriebenen Angriffen stellt der Backscatter-Traffic genau die Pakete da, welche als Antwort auf die vom Angreifer erhaltenen Anfragen versendet werden.

Im Gegensatz zu TCP ist es bei UDP deutlich schwieriger, Antwortpakete als Backscatter-Traffic zu identifizieren. Dies liegt daran, dass UDP, wie in Abschnitt 3.1 beschrieben, nicht verbindungsorientiert ist und somit keine *control-flags* besitzt, um ein Paket als Anfrage oder Antwort zu unterscheiden. Um diese Zuordnung dennoch für UDP-Pakete durchzuführen, wurden in dieser Arbeit die nativen Zeek-Heuristiken verwendet[42]. So ist Zeek in der Lage, Pakete auf Anwendungsebene zu analysieren und somit durch eindeutige Merkmale der Protokolle eine Einteilung zwischen Anfrage und Antwort durch-

zuführen. Bei DNS-Paketen ist eine solche Zuordnung beispielsweise durch das *Query/Response* Bit möglich[14]. Bei NTP-Paketen hingegen kann das *Mode-Feld* für eine solche Klassifizierung verwendet werden[**ntp˙mode˙feld**].

Für diese Arbeit wurden so die Einträge der *history* Spalte aus der conn.log Datei zur Filterung des Backscatter-Traffics verwendet. In dieser Spalte protokolliert Zeek den Kommunikationsverlauf zwischen Sender und Empfänger einer Verbindung. Große Buchstaben stehen hierbei für ein gesendetes Paket des Senders und kleine Buchstaben für die des Empfängers. Zur Erkennung von Backscatter-Verbindungen wird hierfür nach dem Wert *ˆd* gefiltert. Das Zirkumflex-Symbol steht hierbei dafür, dass die Verbindungsrichtung aufgrund von Zeek-Heuristiken getauscht wurde[**zeek˙conn˙protocol**, 42]. Dies bedeutet weiter, dass die Quell- und Ziel-IP vertauscht wurden, da Zeek ein Antwortpaket erkannt hat, ohne vorher hierfür eine Anfrage identifiziert zu haben. Im Falle eines Internet-Teleskops entspricht dies genau der Situation, in welcher das Antwortpaket identifiziert wird, welches die Antwort einer Anfrage ist, dessen src-IP mittels IP-Spoofing erzeugt wurde.

Da dieses Verfahren in den Zeek-Dokumentationen der Heuristiken[42] nicht direkt zur Erkennung von DoS/DDoS-Angriffen aufgeführt ist, wurde die Korrektheit der beschriebenen Vorgehensweise durch eine Simulation bestätigt. Hierzu wurden mittels Python einige Pcap-Dateien erstellt, welche dazu verwendet wurden, Antwortpakete eines Angriffs zu simulieren[43]. Zusätzlich wird das Verfahren durch einen Beitrag der Zeek-Community[44] untermauert, in welchem die genannte Vorgehensweise zur Erkennung von Backscatter-Traffic bestätigt wird.

Falsch-positive Ergebnisse durch Scanning können bei diesem Verfahren vernachlässigt werden, da UDP-Scanner in der Regel keine Antwortpakete versenden.Im Teleskop erhaltene UDP-Antwortpakete, welche auf Fehlkonfigurationen zurückzuführen sind, können hingegen mit der dargestellten Vorgehensweise nicht vom Backscatter-Traffic unterschieden werden. Als *backscatter.log* werden im Folgenden dieser Arbeit alle Verbindungen der conn.log des Cloud-Teleskops bezeichnet, welche als Backscatter-Traffic charakterisiert wurden.

4.6 Geo-Datenbank

Wie an den in Kapitel 5 vorgestellten Ergebnissen zu erkennen, wurden die durch Zeek erzeugten Log-Dateien mit zusätzlichen Geo-Informationen erweitert. Hierzu konnte auf insgesamt sieben verschiedene Geo-Datenbanken der University of Twente zurückgegriffen werden. Diese Datenbanken werden im regelmäßigen Abstand aktualisiert und in einem *MinIO-Object-Storage* als CSV- oder JSON-Datei bereitgestellt. Der Zugriff auf diesen Speicher wurde durch ein Shell-Script[45] umgesetzt, welches jeweils die neuste Version der Datenbanken herunterlädt. Die Informationen der Datenbanken stammen hierbei von dem GeoIP-Datenanbieter IPinfo[46]. Die Datenbanken sind so aufgebaut, dass

jeweils einer IP-Adresse bzw. einem IP-Adressbereich die Geo-Informationen der jeweiligen Datenbank zugeordnet sind.

Um die von Zeek erzeugte conn.log Datei mit diesen Informationen zu erweitern, wurde das Kommandozeilentool *mmdbtctl* verwendet[47]. Hierzu wurden die Datenbanken mittels eines Python-Scripts[48] in das *.mmdb* Format konvertiert. Dieser Dateityp ermöglicht es dem mmdbtctl-Kommandozeilentool, die Informationen einer Geo-Datenbank mit einer Liste an IP-Adressen zu verknüpfen[47]. Dieses Verknüpfungsverfahren wurde mit demselben Python-Script[48] umgesetzt. So wurde das Script einmal für die Quell-IPs und einmal für die Ziel-IPs der conn.log durchlaufen.

Die im weiteren Verlauf der Arbeit am häufigsten verwendeten Informationen stammen aus den Datenbanken *standard-location*, *standard-company* und *rir*. Geografische Informationen wie Kontinent, Land und Stadt wurden hierbei aus standard location entnommen. Aus standard company wurde die Spalte *type* verwendet, welche jeder IP-Adresse einen Wert der vier Typen *hosting*, *business*, *isp* oder *education* zuordnet. In der RIR-Datenbank sind Informationen über die *Regional-Internet-Registry* enthalten. Die Regional-Internet-Registry bezeichnet die regionalen Internet-Registrierungsstellen, welche die jeweiligen IP-Adressbereiche sowie die Nummern der autonomen Systeme (AS) verwalten[49]. Aus dieser Datenbank wurde hauptsächlich die Spalte *as-name* verwendet. Diese Spalte enthält den Namen des Unternehmens bzw. der Organisation, zu welcher das jeweilige AS und somit der IP-Adressbereich zugeordnet ist.

Sowohl für Quell- als auch Ziel-Adressen liegt für alle oben genannten Spalten der Anteil gefundener GEO-Einträge bei über 90%.

5 Ergebnisse

Im Folgenden werden die Analyse-Ergebnisse dieser Arbeit dargestellt. Hierbei werden zunächst einige allgemeine Ergebnisse der vom Cloud-Teleskop aufgezeichneten Daten vorgestellt. Anschließend erfolgt eine vertiefte Betrachtung der identifizierten Scans. Das dritte Unterkapitel betrachtet dieselben Daten des Cloud-Teleskops und beschäftigt sich mit dem aufgezeichneten Backscatter-Traffic. Zum Abschluss erfolgt eine vergleichende Analyse der beiden Teleskope, bei welcher die zuvor im Cloud-Teleskop aufgezeigten Phänomene in den Daten des CAIDA-Teleskops überprüft werden.

5.1 Cloud-Teleskop Allgemein

Alle im weiteren Verlauf dieser Arbeit dargestellten Ergebnisse beziehen sich auf den Analysezeitraum vom 20.01.2025 um 16:00 bis zum 03.02.2025 um 01:33. Dies entspricht einem Zeitraum von ca. zwei Wochen. In diesen zwei Wochen hat das Cloud-Teleskop Pcap-Daten im Umfang von ≈10,6 Gb aufgezeichnet. Nach der Analyse mit Zeek konnten aus diesen Daten ≈20.500.000 Verbindungen extrahiert werden. Alle weiter beschriebenen Analysen sowie das Erzeugen der Abbildungen wurden durch ≈7.800 Zeilen Python-Code umgesetzt. Der komplette Code, welche auch die nachfolgenden Algorithmen enthält, ist in einem GitHub-Repository veröffentlicht[50].

Der prozentuale Anteil des verwendeten Transportprotokolls sämtlicher aufgezeichneter Verbindungen ist im nebenstehenden Kreisdiagramm 5.1 dargestellt. Hierbei sofort festzustellen ist der mit 91,6% überwiegende Anteil an TCP-Verbindungen. Die beiden Protokolle UDP und ICMP sind hingegen jeweils für etwa 4% der Verbindungen verantwortlich.

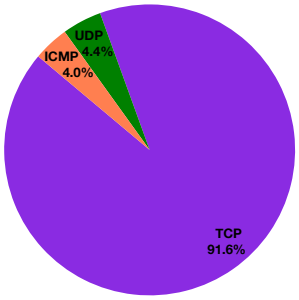


Abbildung 5.1: Prozentuale Verteilung der Transportprotokolle.

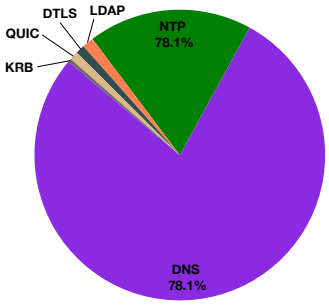


Abbildung 5.2: Prozentuale Verteilung gefundener Anwendungsprotokolle.

Genauer aufgeschlüsselt werden diese UDP-Verbindungen in 5.2. Das Kreisdiagramm visualisiert die prozentuale Verteilung der zu diesen Verbindungen gefundenen Protokolle der Anwendungsschicht. Die Zuordnung der Verbindungen wurde, wie in Abschnitt 4.1 beschrieben, mittels Zeek umgesetzt. Auffallend hierbei ist der große Anteil an DNS-Verbindungen, welcher fast 80% ausmacht. Mit ca. 1/5 der im Diagramm betrachteten Verbindungen steht das Network-Time-Protocol für den zweitgrößten Anteil. Zusätzlich sind eine Reihe von Protokollen zu erkennen, welche allerdings nur einen kleinen Anteil zu verantworten haben.

5.2 Scanning

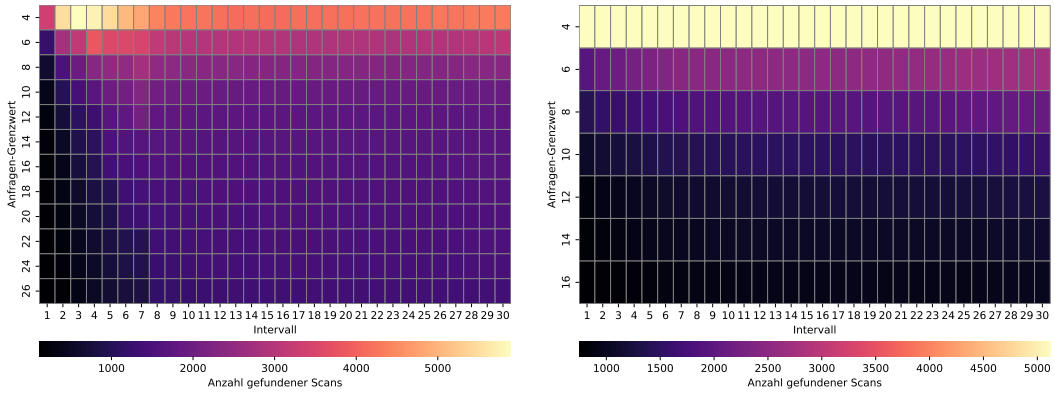
Die Ergebnisse zu dem Thema Scanning werden im Anschluss in vier Unterkapiteln dargestellt. Zu Beginn erfolgt eine Betrachtung des Scan-Verhaltens bei unterschiedlichen Werten für das zeitliche Intervall sowie für den Anfragen-Grenzwert. Anschließend findet eine zeitliche und geografische Betrachtung der identifizierten Port- und Address-Scans statt. Abschließend folgt eine genauere Betrachtung der auffälligsten Scans, sowie eine Auflistung der in den Scans gefundenen Forschungseinrichtungen.

5.2.1 Scan-Klassifizierung anhand des zeitlichen Intervalls und der Anzahl registrierter Verbindungen

Wie in Abschnitt 4.4 bereits erläutert, ist zur Identifizierung von Scans die Auswahl eines zeitlichen Intervalls sowie eines minimalen Grenzwertes an Verbindungen von entscheidender Bedeutung. Diese Parameter sind maßgeblich dafür verantwortlich, zu entscheiden, welche src-IPs bei welcher Anzahl an Anfragen über welchen Zeitraum hinweg als Scans erkannt werden oder nicht.

Die beiden Heatmaps aus 5.3 visualisieren die Veränderung der identifizierten Scans in Abhängigkeit eines steigenden Intervalls sowie eines steigenden Anfragen-Grenzwertes. Die x-Werte repräsentieren hierbei jeweils das zeitliche Intervall in Minuten und die y-Werte den Anfragen-Grenzwert. Heatmap 5.3a veranschaulicht die gefundenen Address-Scans und 5.3b bezieht sich auf Port-Scans. Zur Erstellung der beiden Diagramme wurde der in Abschnitt 4.4 beschriebene Algorithmus in wiederholter Weise für die Anfragen-Grenzwerte von 4 bis 26 und die Intervallwerte von 1 Min bis 30 Min durchgeführt. Hierzu wurde der Algorithmus so erweitert, dass neben dem Intervallwert auch der Anfragen-Grenzwert als Parameter gesetzt wird. Somit erfolgt in jedem Iterationsschritt vor der Prüfung des Intervallwertes eine Überprüfung des Anfragen-Grenzwertes. Die zuvor von Zeek gefundenen src-IPs können weiterhin als Vorfilterung verwendet werden. Dies ist damit zu begründen, da das beschriebene Zeek-Script mit einem Anfragen-Grenzwert von 4 durchgeführt wurde. Da dies genau der kleinste x-Wert der Heatmaps ist, gehen durch die

Vorfilterung keine Scans verloren. Da für die Port-Scans in 5.3b oberhalb des Anfragen-Grenzwertes von 16 keine Veränderungen zu erkennen sind, wurde die Heatmap bei diesem y-Wert limitiert. Grundlegend bei beiden Heatmaps zu erkennen ist, dass die Anzahl identifizierter Scans bei steigendem Anfragen-Grenzwert abnimmt. Für die Port-Scans gilt, mit Ausnahme des Intervallswertes von 4, dass die erkannten Scans bei steigendem Intervall zunehmen. Anders hingegen ist es bei den Address-Scans, hier lässt sich vor allem bei einem niedrigen Anfragen-Grenzwert ein anderes Phänomen beobachten. So wächst die Anzahl gefundener Scans zunächst bei steigendem Intervall. Dies geschieht aber nur bis zu einer bestimmten Intervallgrenze. Nach Überschreitung dieser Grenze nimmt die Anzahl der Scans wieder ab. Daran anschließend zu beobachten ist zudem, dass sich diese Intervallgrenze bei steigendem Anfragen-Grenzwert erhöht. So liegt bei einem y-Wert von 6 diese Grenze bei dem Intervallwert von 6. Bei einem y-Wert von 10 liegt diese Intervallgrenze hingegen bei einem Wert von 8.



(a) Address-Scans **(b)** Port-Scans

Abbildung 5.3: Anzahl gefundener Scans in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls.

Da die beiden zuvor beschriebenen Heatmaps alleine nicht ausreichen, um eine begründete Entscheidung der beiden Parameter zu fällen, wurden für diese Arbeit zusätzlich die folgenden Heatmaps in 5.4 sowie 5.5 erstellt. Alle diese Abbildungen sind ähnlich zu obigen aufgebaut und beschreiben einen Verlauf in Abhängigkeit des Intervalls sowie des Anfragen-Grenzwertes.

Die beiden Heatmaps aus 5.4 beschreiben hierbei jeweils für Address- und Port-Scans die Entwicklung neu hinzukommender Scanner. Zur Erstellung dieser Heatmaps wurde der oben genannte Algorithmus ein weiteres Mal angepasst und erneut für entsprechende x- und y-Werte durchlaufen. Entscheidend, um als neuer Scanner kategorisiert zu werden, ist die src-IP des gefundenen Scans. So werden für jeden Anfragen-Grenzwert nur die Scanner zum jeweiligen Intervall hinzugezählt, wo gilt, dass die src-IP des gefundenen Scans in keinem vorherigen Intervall vorgekommen ist. Der Algorithmus durchläuft somit

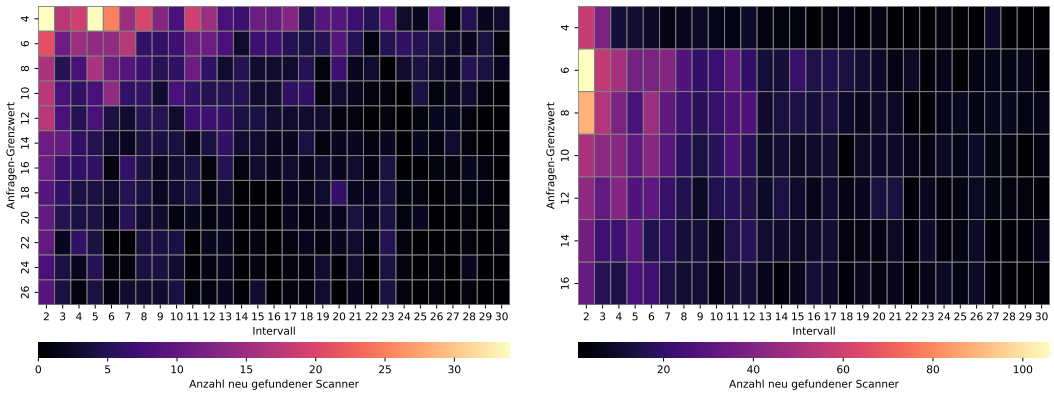
weiterhin alle Anfragen-Grenzwerte und Intervalle um mögliche Scans zu identifizieren, überprüft nun bei gefundenen Scans zunächst, ob die src-IP schon in einem vorherigen Scan des Anfragen-Grenzwertes vorgekommen ist.

Die x-Werte der beiden Heatmaps beginnen jeweils bei Minute 2, dies ist dem geschuldet, dass die Werte bei einem Intervall von einer Minute überdurchschnittlich groß sind und so die Farbskala unproportional beeinflussen würden. Dies lässt sich damit begründen, dass der Algorithmus beim Startintervall über keine zuvor gefundenen Scanner verfügt, mit welchen ein Abgleich der in Minute eins gefundenen Scanner möglich ist.

Zu den Address-Scannern aus 5.4a ist zu beobachten, dass die Dichte der neu gefundenen Scanner bei steigendem Anfragen-Grenzwert abnimmt. Mit Ausnahme des Anfragen-Grenzwertes von 4 lässt sich diese Beobachtung auch auf die Port-Scanner übertragen. Dieser x-Wert weicht insofern von dem Muster ab, da bei gleichem Intervallwert die Anzahl der neu gefundenen Scanner kleiner ist als die der Anfragen-Grenzwerte von 6 und 8.

Eine solche monotone Abnahme lässt sich bei steigendem Intervall in keinem der beiden Diagramme feststellen. Für die Address-Scanner ist beispielsweise am dem Anfragen-Grenzwert von 4 zu erkennen, dass die Anzahl der neu gefundenen Scanner für den Intervallwert von 3 kleiner ist, als die des Intervallwertes von 4. Für die Port-Scanner ist Gleiches für den Anfragen-Grenzwert von 8 und den Intervallwerten von 5 und 6 festzustellen.

Dennoch kann für beide Scan-Typen bei Betrachtung der gesamten Heatmap eine Tendenz für einen solchen Rückgang beobachtet werden. So ist für alle Anfragen-Grenzwerte eine Intervallgrenze zu erkennen, ab welcher keine neuen Scanner mehr gefunden werden. Weiter in diesem Zusammenhang zu beobachten ist, dass diese Grenze bei steigendem Anfragen-Grenzwert kleiner ausfällt. Hierbei anzumerken ist, dass dies nur für einen Großteil der Werte gilt und vereinzelt Ausreißer zu verzeichnen sind. Eine solche Abweichung ist beispielsweise für die Address-Scanner bei dem Intervallwert von 23 zu erkennen.

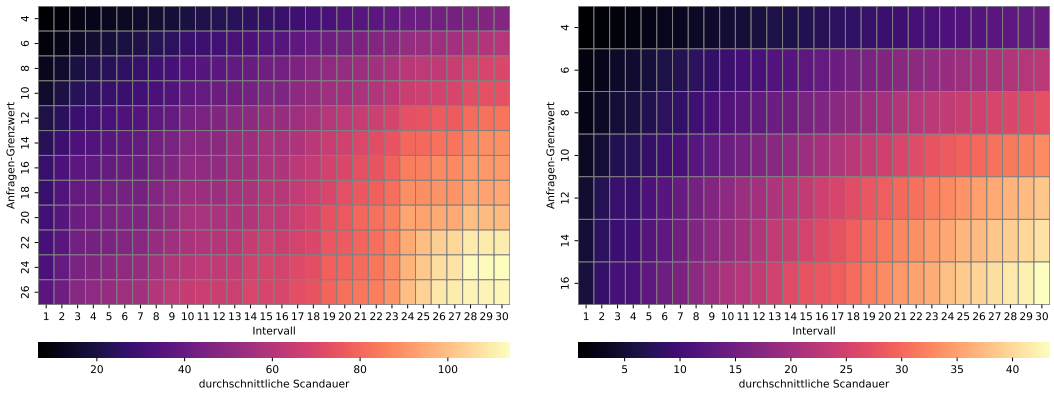


(a) Address-Scanner **(b)** Port-Scanner

Abbildung 5.4: Anzahl neu gefundener Scanner in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls.

Weitere interessante Aspekte zum Verständnis des Scan-Verhaltens lassen sich durch die Betrachtung der durchschnittlichen Scandauer identifizieren. Zur Erstellung der Heatmaps aus 5.5 wurde der oben genannte Code ein letztes Mal angepasst. Für alle gefundenen Scans werden nun die genauen Zeitstempel der zugehörigen Anfragen gespeichert. Somit kann für jeden Scan aus der Differenz der zeitlich zuletzt und zuerst registrierten Anfrage die Scandauer ermittelt werden. Der Durchschnitt der berechneten Werte aller Scans eines bestimmten Intervalls und eines bestimmten Anfragen-Grenzwertes ergibt den jeweiligen Wert im Diagramm. Wie schon zuvor wurde diese Berechnung jeweils für Address-Scans 5.5a als auch Port-Scans 5.5b durchgeführt. Die Werte der Farbskala repräsentieren hierbei jeweils die Scandauer in Minuten.

Auffällig bei Betrachtung der Diagramme ist, dass der Verlauf der Scandauer beider Heatmaps nahezu identisch ist. So gilt, dass bei steigendem Anfragen-Grenzwert sowie bei steigendem Intervall die Scandauer zunimmt. Große Unterschiede der beiden Diagramme liegen allerdings in der Skalierung der Farbwerte.



(a) Address-Scans **(b)** Port-Scans

Abbildung 5.5: Durchschnittliche Scandauer in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls.

Wie im Kapitel 6 nachzulesen, wurde für das Cloud-Teleskop der Anfragen-Grenzwert zur Erkennung von Port- und Address-Scans auf 6 festgelegt. Um für diesen Anfragen-Grenzwert eine begründete Entscheidung für ein entsprechendes Intervall vornehmen zu können, werden die Erkenntnisse der Graphen aus 5.6 herangezogen. Hierbei bezieht sich 5.6a auf die Address- und 5.6b auf Port-Scans.

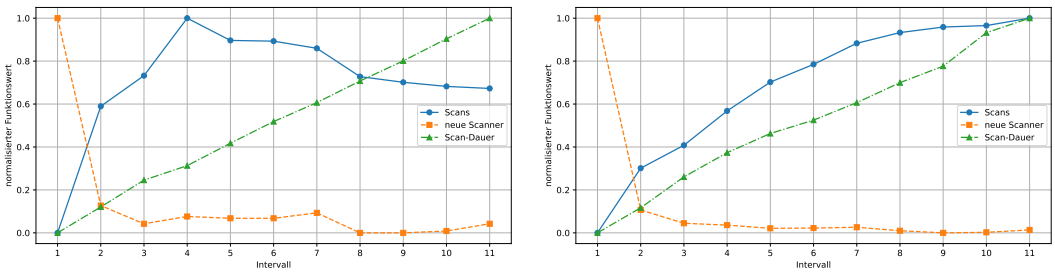
In den Koordinatensystemen sind für den festgelegten Anfragen-Grenzwert von 6 die Werte aus den obigen Heatmaps 5.3, 5.4 und 5.5 visualisiert. Somit skizzieren die Diagramme einen Vergleich zwischen der Anzahl gefundener Scans, der Anzahl neu gefundener src-IPs sowie der durchschnittlichen Scandauer. Bei diesem Vergleich interessieren weniger die totalen Werte als als eine vergleichbare Darstellung der drei Datensätze. Aus diesem Grund wurden die

zur Erzeugung der Diagramme genutzten Werte zuvor normalisiert. Die normalisierten Werte wurden hierbei durch die unterstehende Formel bestimmt. Der Parameter X_{min} steht hierbei für den geringsten Wert des jeweiligen Datensatzes und X_{max} für den höchsten Wert.

$$X_{\text{normalisiert}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Somit ergeben sich für die y-Achse Werte zwischen 0 und 1. Auf der x-Achse sind jeweils die Intervalle von 1 bis 10 abgebildet. Das Intervall von 10 wurde hier als Grenze gesetzt, da eine höhere Intervallgrenze keine relevanten Befunde liefert.

Der Funktionsverlauf der identifizierten Address-Scans in 5.6a entspricht dem in 5.3a erkannten Muster. So steigt der Graph zunächst bis zu einem Intervallwert von 4. Ab diesem Intervallwert ist ein Rückgang zu verzeichnen, wobei die Abnahme zwischen den Intervallwerten von 4 und 5 sowie von 7 und 8 am größten ausfällt. Bei der Anzahl neu gefundener src-IPs ist zu beobachten, dass diese zwischen den Intervallwerten von 2 bis 7 nahezu konstante Funktionswerte aufweisen. Zwischen den Intervallwerten von 7 und 8 erfolgt dann eine ruckartige Dekrementierung. Anschließend bewegen sich die Werte um den Nullpunkt. Der Funktionsverlauf der durchschnittlichen Scandauer ist nahezu identisch wie der der Port-Scans in 5.6b. Beide Funktionsverläufe weisen hier eine monotone und nahezu gleichmäßige Steigung auf. Dies hat zu bedeuteten, dass durch Erhöhung des zeitlichen Intervalls auch die Scandauer um einen konstanten Faktor steigt. Neben der Scandauer weist bei den Port-Scans auch die Anzahl der gefundenen Scans eine monotone Steigung auf. Hierbei ist für kleinere Intervallwerte ein größerer Anstieg festzustellen als bei höheren Werten. Zu der Anzahl neu gefundener src-IPs aus 5.6b sind zwischen den Intervallwerten von 4 bis 7 nahezu konstante Funktionswerte festzustellen. Ab diesem Intervallwert ist ein Rückgang zu verzeichnen, welcher bei dem Intervallwert von 10 im Nullpunkt endet.



(a) Address-Scans **(b)** Port-Scans

Abbildung 5.6: Normalisierte Funktionsverläufe der Anzahl gefundener Scans, der Anzahl neu gefundener Scanner und der durchschnittlichen Scandauer bei einem Anfragen-Grenzwert von 6.

Wie in Kapitel 6 unter Verwendung der zuvor beschriebenen Koordinatensysteme hergeleitet, ist für alle folgenden Analysen das Intervall zur Erkennung von Address-Scans auf 6 Minuten und zur Erkennung von Port-Scans auf 7 Minuten festgelegt.

5.2.2 Zeitliche Betrachtung des Scan-Verhaltens

Um einen genaueren Überblick der im Analysezeitraum durchgeführten Scans zu erhalten, ist in 5.7 die zeitliche Entwicklung der vom Cloud-Teleskop identifizierten Scans visualisiert. Zur Veranschaulichung der Unterschiede zwischen Address- und Port-Scans ist in dem Koordinatensystem für beide Typen ein Verlauf der entdeckten Scans dargestellt. Um die Lesbarkeit des Diagramms zu erhöhen, wurden die gefundenen Scans in 60-Minuten-Intervalle aggregiert. Um zu verhindern, dass einzelne Ausreißer die grafische Darstellung verzerren, kommt ab einem y-Wert von 20 eine logarithmische Skalierung zum Einsatz. Die x-Achse entspricht den Tagen des Analysezeitraums (20.01.2025 16:00 - 03.02.2025 01:33).

Zur Erstellung der Koordinatensysteme wurde ein entsprechender Python-Algorithmus entworfen. Dieser verwendet die Daten aus address-scan.log und port-scan.log, um die Ausgangsdaten der Abbildungen zu erzeugen. Hierzu wird dem Algorithmus eine Zeiteinheit (Sekunde, Minute, Stunde, etc.) sowie ein Start- und End-Zeitstempel mitgegeben. Die grundlegende Funktionsweise besteht darin, den gesetzten Zeitraum zu durchlaufen und alle zu einer Zeiteinheit gefundenen Scan-Einträge aufzuaddieren. Wird zu einer Zeiteinheit kein Eintrag gefunden, so wird für diese eine Null gesetzt.

Am auffälligsten bei der Betrachtung des Diagramms ist, dass der Graph der gefundenen Address-Scans fast ausschließlich oberhalb dem der Port-Scans verläuft. So bewegt sich die Anzahl gefundener Address-Scans überwiegend zwischen den Werten von 8 bis 20, wohingegen die Werte der Port-Scans hauptsächlich zwischen 0 und 8 verlaufen. Weiter lassen sich drei deutliche Ausreißer identifizieren. Um herauszufinden, welche Scans diese Ausreißer zu verantworten haben, wurden die entsprechenden Zeitabschnitte genauer untersucht. Hierzu wurde die address-scan.log bzw. port-scan.log nach diesen Zeitabschnitten gefiltert. Die gefilterten Daten wurden anschließend anhand ihrer src-IP gruppiert und so die Anzahl der Vorkommen jeder IP bestimmt. Die zugehörigen RIR-Namen wurden aus der conn.log entnommen.

Durch dieses Verfahren ist zu erkennen, dass aus dem IP-Bereich von *ssd networks limited* am 24.01 zwischen 20 und 22 Uhr eine enorme Anzahl an Address-Scans durchgeführt wurden. Die aus *ssd networks limited* durchgeführten Scans entsprechen hierbei 96,9% aller in diesem Zeitraum registrierten Scans. Hingegen sorgten am 26.01 *Mobile Communication Company of Iran PLC* und *Iran Telecommunication Company PJS*, mit zusammengerechnet 65%, für ein erhebliches Aufkommen an Port-Scans. Der letzte Ausreißer am 30.01 zwischen 10 und 15 Uhr lässt sich auf die RIR-Namen *Information*

Technology Company (ITC), *Saudi Telecom Company JSC* und *Mobile Communication Company of Iran PLC* zurückführen.

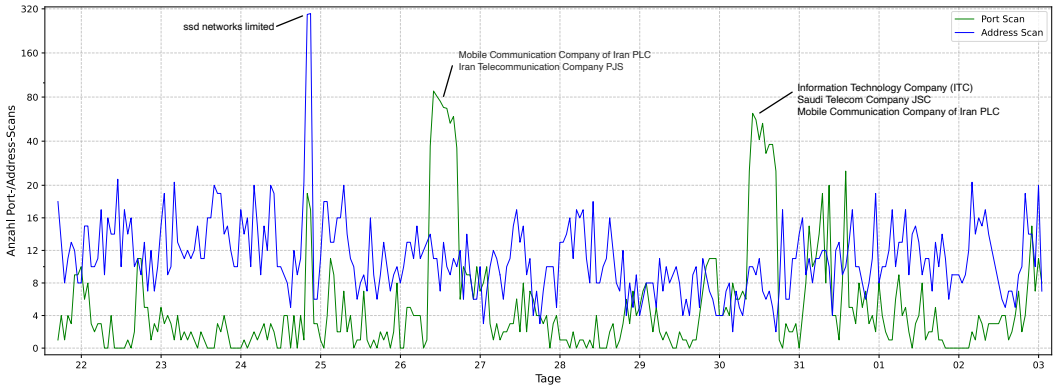


Abbildung 5.7: Zeitlicher Verlauf identifizierter Port- und Address-Scans (60-Minuten-Aggregation).

Um die zuvor beobachteten Phänomene der identifizierten Scans mit den tatsächlich registrierten Scan-Anfragen vergleichen zu können, wurde die Abbildung 5.8 erstellt. Das Koordinatensystem ist hierbei ähnlich aufgebaut wie das aus 5.7. Der y-Wert zum Übergang auf eine logarithmische Skalierung ist hierbei auf 900 gesetzt. Zur Ermittlung der Ausgangsdaten wurde der zuvor beschriebene Algorithmus verwendet. Um hierbei den zeitlichen Verlauf der tatsächlichen Scan-Anfragen zu erhalten, wurden die Einträge der conn.log anhand der Scans aus address-scan.log bzw. port-scan.log gefiltert.

Ähnlich wie in 5.7 ist zu erkennen, dass der Graph der Address-Scan-Anfragen bis auf wenige Ausnahmen oberhalb dem der Port-Scan-Anfragen verläuft. Andere sich wiederholende Muster sind für keinen der beiden Graphen zu erkennen. Um mehr Informationen über die im Diagramm visualisierten Hochpunkte zu erhalten, wurden die entsprechenden Zeiträume in der zuvor gefilterten conn.log nach der oben beschriebenen Vorgehensweise genauer überprüft. Für alle aufgeführten RIR-Namen wurden zusätzlich DNS-Reverse Anfragen der zugehörigen IPs durchgeführt. Haben diese Anfragen nicht zu sinnvollen Ergebnissen geführt, so wurde der entsprechende RIR-Name als Beschriftung verwendet. Für den RIR-Namen *DigitalOcean, LLC* wurde zusätzlich die für den Ausreißer verantwortliche IP angegeben.

Zunächst festzustellen ist, dass alle in 5.7 identifizierten Ausreißer auch bei Betrachtung der tatsächlichen Scan-Anfragen zu erkennen sind. Für diese Ausreißer gilt zudem, dass die gleichen RIR-Namen auch in den entsprechenden Zeiträumen aus 5.8 vorzufinden sind. Ein Unterschied der beiden Diagramme besteht darin, dass für die drei genannten Zeiträume in 5.8 jeweils ein Hochpunkt für beide Scan-Typen vorzufinden ist. Bemerkenswert ist allerdings, dass für die beiden Zeiträume am 26.02 und 30.01 gilt, dass der RIR-Name der Address-Scans mit *DigitalOcean, LLC (188.166.250.148)* nicht derselbe ist wie der der Port-Scans.

Für den Ausreißer am 24.01 ist dieses Phänomen hingegen nicht vorzufinden. Hier sorgen Anfragen des RIR-Namens *ssd networks limited* auch bei den Port-Scans für 92% aller in diesem Zeitraum registrierten Scan-Anfragen. Ein weiterer Unterschied der beiden Diagramme ist in der Zeitspanne zwischen dem 28.01 um 19 Uhr bis zum 29.01 um 2 Uhr zu beobachten. In diesem Zeitraum ist in 5.8 eine extreme Anzahl an Port-Scan-Anfragen zu verzeichnen. Dieser Hochpunkt ist allerdings weder in dem Verlauf der Port-Scans noch in dem der Address-Scans in 5.7 festzustellen. Dies bedeutet wiederum, dass von *gcommer.com* mittels weniger Scans eine extreme Last an Anfragen erzeugt wurde.

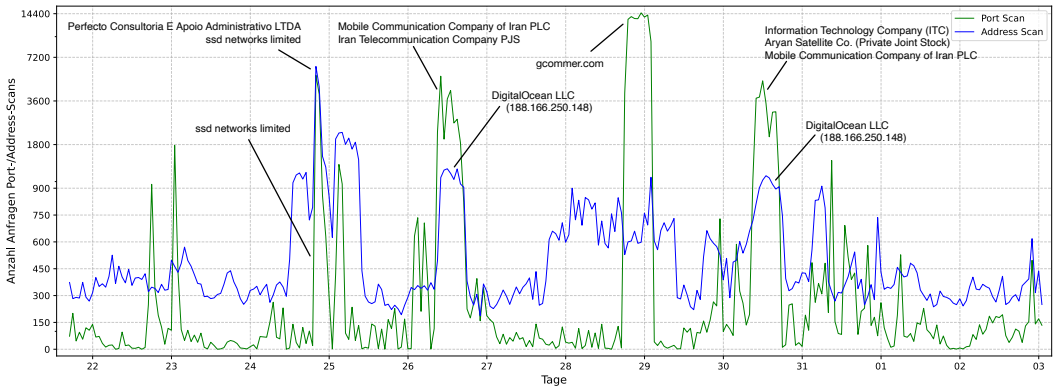


Abbildung 5.8: Zeitlicher Verlauf registrierter Anfragen der Port- und Address-Scans (60-Minuten-Aggregation).

5.2.3 Geografische Betrachtung des Scan-Verhaltens

Um weitere Informationen über die identifizierten Scans zu erhalten, wird im Folgenden eine geografische Betrachtung des Scan-Ursprungs durchgeführt. So sind in 5.9 die Top 20 Länder aufgeführt, welche für die meisten Scans verantwortlich sind. Zu jedem der Länder ist hierbei jeweils die Anzahl an Address-Scans, Port-Scans und Mixed-Scans angegeben.

Wie in Abschnitt 3.6 beschrieben, ist ein Mixed-Scan die Kombination aus Address- und Port-Scan. Also ein Scan, welcher auf mehreren Ziel-Hosts mehrere Ports anfragt. Zur Berechnung dieser Mixed-Scans wurden jeweils alle Scans aus address-scan.log und port-scan.log anhand gleicher src-IPs in einzelne Gruppen eingeteilt. Somit entsteht zu jeder src-IP eine zugehörige Address-Scan-Gruppe und eine zugehörige Port-Scan-Gruppe. Diese Gruppen wurden anschließend miteinander verglichen. Bei diesem Vergleich wurden zunächst alle Scans der Address-Scan-Gruppe durchlaufen und jeweils der Zeitraum zwischen Start- und Endzeitpunkt des Scans ermittelt. Eine Klassifizierung als Mixed-Scan erfolgt, wenn in der Port-Scan-Gruppe der gleichen src-IP ein Scan enthalten ist, dessen Startzeitpunkt innerhalb dieses Zeitraums liegt. Nach Durchlaufen aller Address-Scan-Gruppen wurden die gefundenen Mixed-Scans

aus beiden Gruppen entfernt. Der gleiche Vergleich wurde nun in umgekehrter Weise für die Scans der Port-Scan-Gruppe durchgeführt. Um anschließend die Kategorisierung anhand der Länder vorzunehmen, wurden die src-IPs mit den in der conn.log enthaltenen Geo-Informationen abgeglichen.

Bei Betrachtung des Diagramms ist sofort zu erkennen, dass ein kleiner Teil der Länder für den Großteil der Scans verantwortlich ist. So verursachen die Top fünf Länder zusammen 84,6% des gesamten Scan-Aufkommens. Weiter auffällig ist die ungleichmäßige Verteilung der Scan-Typen auf die einzelnen Länder. Unter Betrachtung der Top 5 Länder sind es hier die USA, Deutschland, Großbritannien und die Niederlande von denen hauptsächlich Address-Scans durchgeführt werden. Aus dem Iran werden hingegen ausschließlich Port-Scans durchgeführt. Eine weitere Besonderheit ist an den Mixed-Scans zu erkennen. Diese werden ausschließlich von den Ländern USA, Großbritannien, Niederlande und China durchgeführt. Zudem festzustellen ist, dass unter den Top 20 Ländern die Anzahl der Address-Scans mit 67,6% mehr als doppelt so hoch ist wie die der Port-Scans (32,3%).

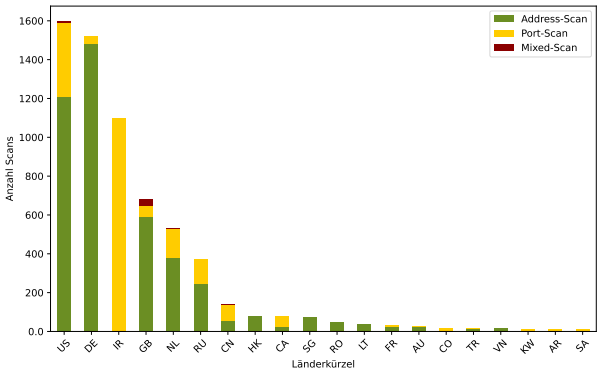


Abbildung 5.9: Verteilung der identifizierten Scans auf die Top 20 Herkunftsländer sowie die Einteilung anhand der Scan-Typen.

Um weiterführende Informationen über das Scan-Verhalten der einzelnen Länder zu erhalten, hilft eine Betrachtung der von diesen Ländern gescannten Ports. Hierzu folgt mit 5.10 zunächst ein allgemeiner Überblick über die von allen identifizierten Address-Scans angefragten Ports. Auf der x-Achse sind die Top 20 Ports aufgeführt, welche am häufigsten von Address-Scans verwendet wurden. Die Ports sind hierbei zu den Protokollnamen aufgelöst. Die y-Achse steht für den Prozentwert des jeweiligen Protokolls. Der zu einem Address-Scan zugehörige Port konnte aus den Einträgen der *address-scan.log* entnommen werden. Dieser Wert wird im angewandten Algorithmus bei der Erstellung des jeweiligen Eintrages

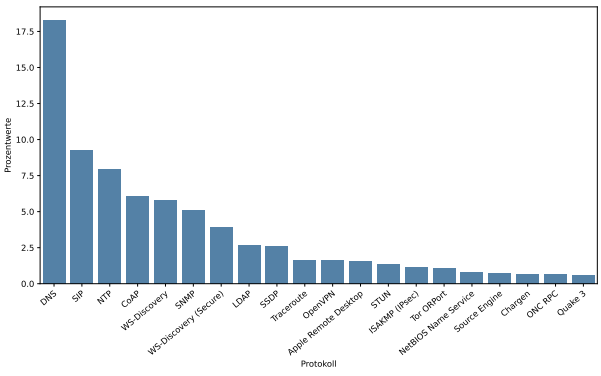


Abbildung 5.10: Prozentuale Verteilung der Top 20 am meisten gescannten Protokolle.

gesetzt.

Bei Betrachtung des Diagramms ist schnell zu erkennen, dass Port 53 (DNS) mit über 18% fast ein Fünftel aller gescannten Protokolle ausmacht. An zweiter Stelle ist das unter Port 5060 laufende *Session Initiation Protocol* (SIP) zu finden. Dieses Protokoll wird hauptsächlich für Voice-over-IP Anwendungen verwendet und dient hier zur Steuerung, Einrichtung, Verwaltung und Beendigung von Kommunikationssitzungen[10]. Unter den Top 5 sind in dem Diagramm zudem die Protokolle *NTP*, *Constrained Application Protocol* sowie *WS-Discovery* zu finden. Das *Constrained Application Protocol* (Port 5683) ist ein Web-Transfer-Protokoll, welches speziell für ressourcenbeschränkte Internet-of-Things Geräte entwickelt wurde[51]. Das *Web Services Discovery Protokoll* wird hingegen zur automatischen Erkennung von Webservices in einem lokalen Netzwerk verwendet. So können Geräte ohne manuelle Konfiguration verfügbare Dienste finden und mit diesen kommunizieren[52].

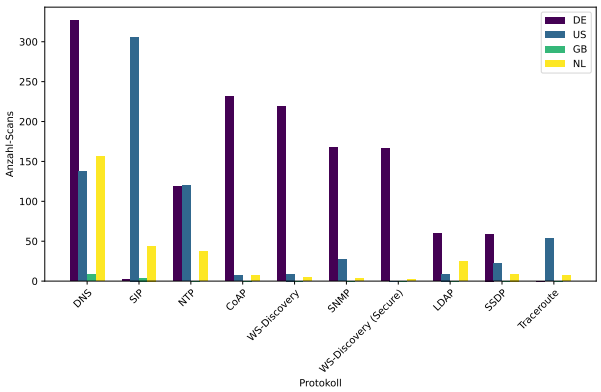


Abbildung 5.11: Verteilung der Scans auf die Top 10 am meisten gescannten Protokolle, differenziert nach den Scan-Herkunfsländern.

Nach der allgemeinen Betrachtung der Protokolle wird nun in 5.11 ein Zusammenhang zu den zugehörigen Ländern hergestellt. Die x-Achse enthält hierbei die Top 10 Protokolle des vorherigen Diagramms 5.10. Zu jedem dieser Protokolle sind vier Säulen zugeordnet. Diese Säulen stehen für die Top vier Länder, welche die meisten Address-Scans zu verzeichnen haben (Deutschland, USA, Großbritannien und Niederlande). Die Höhe der Säulen entspricht der Anzahl, wie viele Address-Scans des jeweiligen Landes das entsprechende Protokoll verwendet haben. Zur Errechnung dieser Werte wurden die Ausgangsdaten der beiden vorherigen Diagramme 5.9 und 5.11 verwendet. Auffällig bei Betrachtung des Diagramms ist das sehr unterschiedliche Scan-Verhalten der einzelnen Länder. So werden von Deutschland aus eine vergleichsweise hohe Anzahl an Address-Scans auf fast alle im Diagramm enthaltenen Protokolle durchgeführt. Für die USA lässt sich sagen, dass diese hauptsächlich für Scans auf das *Session Initiation Protocol* Protokoll verantwortlich sind. Obwohl Großbritannien die drittgrößte Anzahl an Address-Scans zu verantworten hat, sind die Säulen dieses Landes im Diagramm nur schwach ausgeprägt. Zu dem Scan-Verhalten der Niederlande lässt sich kein konkretes Muster feststellen.

5.2.4 Genauere Betrachtung des Scan-Ursprungs

Nachfolgend wird eine genauere Betrachtung des Scan-Ursprungs anhand der zugehörigen RIR-Namen durchgeführt. Hierbei interessiert vor allem der Vergleich zwischen den gefundenen RIR-Namen anhand einzelner Scan-Charakteristika. Für diesen Vergleich werden in 5.12 verschiedene Informationen zu den Top 10 RIR-Namen bereitgestellt, welche für die größte Anzahl an Address-Scans verantwortlich sind. So ist für jeden der RIR-Namen die Anzahl an Scans, die Summe der Anfragen dieser Scans, die durchschnittlichen Scandauer und die Anzahl unterschiedlich gescannt Ziel-Ports visualisiert. Wichtig hierbei zu beachten ist, dass die Skalierung der Anzahl der Scan-Anfragen ab dem Wert von 10.000 eine logarithmische Steigung aufweist. Gleiches gilt für die Anzahl unterschiedlicher Ziel-Ports, hier ist eine logarithmische Skalierung ab dem Wert von 10 vorzufinden.

Für den RIR-Namen *DigitalOcean, LLC* wurden die zugehörigen src-IPs genauer betrachtet. Hier sind die IPs 174.138.84.177, 188.166.72.112, 188.-166.250.148 und 137.184.3.191 für den größten Teil der Scans verantwortlich. DNS-Reverse-Anfragen ergaben ausschließlich zu 137.184.3.191 den Treffer *gorillatacopf.com*. Dies ist ein in den USA ansässiges mexikanisches Restaurant. Bei der weiteren Betrachtung des Diagramms und dem Vergleich der einzelnen RIR-Namen anhand der vier Parameter, sind deutliche Unterschiede in dem Scan-Verhalten zu erkennen. So ist festzustellen, dass *ssd networks limited* und *Arbpr Networks, Inc* für Address-Scans verantwortlich sind, welche an viele verschiedene Ziel-Ports gerichtet sind. Weiter zu beobachten ist, dass von *Pfcloud UG* und *DigitalOcean, LLC* sehr große Scans ausgehen, welche jeweils eine enorme Anzahl an Anfragen versenden. Dies ist an der großen Anzahl an Scan-Anfragen bei einer vergleichsweise geringen Anzahl an Scans festzustellen. Für diese beiden sowie für *ssd networks limited* ist zudem erkennbar, dass diese sehr lange Scans (> 33min) durchführen.

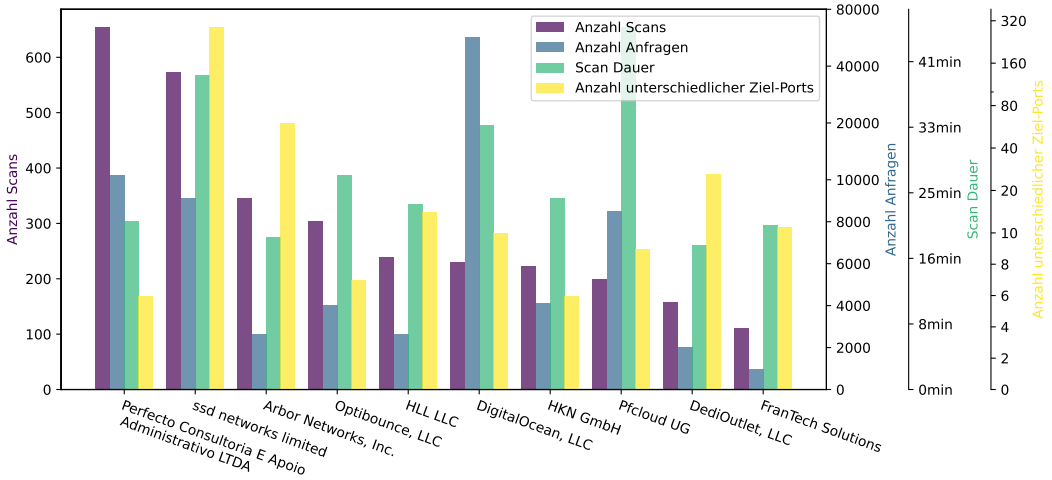


Abbildung 5.12: Address-Scans: Vergleich der Top zehn RIR-Namen anhand der Anzahl an Scans, der Anzahl an Scan-Anfragen, der durchschnittlichen Scandauer und der Anzhal unterschiedlich gescannter Ziel-Ports.

Um oben genannte Beobachtungen auch für Port-Scans zu überprüfen und um einen Vergleich dieser beiden Scan-Typen zu erhalten, wurde das Diagramm 5.13 erstellt. Da dieses Diagramm Port-Scans betrachtet, wurde als vierter Vergleichsparameter die Anzahl der unterschiedlich gescannten Ziel-IPs verwendet. Die Skalierung dieses Parameters wechselt ab dem Wert von 50 auf eine logarithmische Skalierung. Für die Anzahl an Scan-Anfragen gilt dies ab einem Wert von 1000. Da bei den registrierten Port-Scans eine enorme Varianz in Bezug auf die Scandauer vorzufinden ist, beginnt hier die Skalierung im Bereich der Millisekunden und wechselt anschließend mit einer logarithmischen Steigung in den Bereich der Sekunden und Minuten.

Bei Betrachtung des Diagramms ist zunächst festzustellen, dass *Mobile Communication Company of Iran PLC* mit 657 Scans im Vergleich zu den anderen RIR-Namen für einen sehr großen Anteil (41%) aller Scans verantwortlich ist. Weiter zu beobachten ist, dass diese Scans, wie auch die von *Iran Telecommunication Company PJS*, nur auf eine sehr geringe Anzahl an Ziel-IPs abzielen. Zu der Anzahl an Scan-Anfragen lässt sich sagen, dass vor allem von *Information Technology Company (ITC)*, *Rethem Hosting LLC* und *ssd networks limited* sehr große Scans ausgehen. Scans mit einer sehr langen Dauer werden hingegen von *AntiDDoS Solutions LLC*, *CHINANET-BACKBONE* und *NSEC - Sistemas Informaticos, S.A.* durchgeführt.

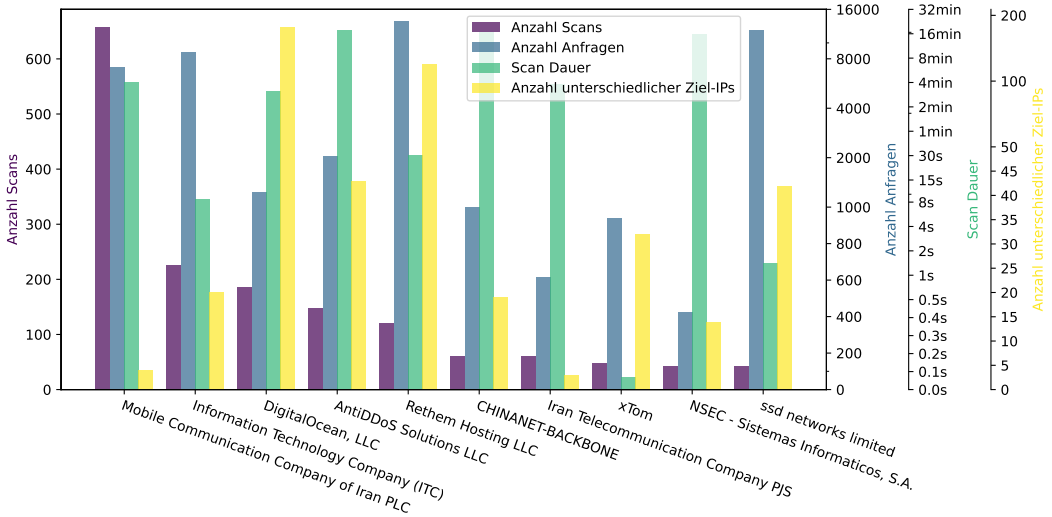


Abbildung 5.13: Port-Scans: Vergleich der Top zehn RIR-Namen anhand der Anzahl an Scans, der Anzahl an Scan-Anfragen, der durchschnittlichen Scandauer und der Anzhal unterschiedlich gescannter Ziel-IPs.

Zum Abschluss dieses Kapitels erfolgt nun mit 5.14 eine Betrachtung der zu Forschungszwecken betriebenen Scans. Zur Charakterisierung dieser Scans wurde die in Abschnitt 4.6 genauer erläuterte *type*-Spalte der *standard-company* Geo-Datenbank verwendet. Die Filterung der Scans erfolgte hierbei auf den *education*-Eintrag. Um die verantwortlichen Forschungseinrichtungen zu iden-

tifizieren, wurden die gefilterten Scans mit den zugehörigen RIR-Namen und den zugehörigen Länderkürzeln erweitert. Zusätzlich wurde DNS-Reverse-Anfragen für alle src-IPs durchgeführt. Um anschließend einen tieferen Einblick in die Forschungsabsichten zu erhalten, wurden die von den Forschungseinrichtungen gescannten Ports ermittelt. Zu beachten ist, dass zu *China Education and Research Network Center* aufgrund fehlender DNS-Reverse-Einträge keine konkrete Forschungseinrichtung ermittelt werden konnte. Bei der *Technische Hochschule Georgia* war der RIR-Name für diese Zuordnung ausschlaggebend genug.

Weiter anzumerken ist, dass die gefundenen Forschungseinrichtungen ausschließlich Address-Scans durchgeführt haben. Eine Besonderheit lässt sich bei dem Vergleich der RIR-Namen feststellen. So ist zu beobachten, dass viele Forschungseinrichtungen ein nationales Forschungsnetz verwenden, um entsprechende Scans durchzuführen. In Deutschland ist dies das Forschungsnetz des *Vereins zur Foerderung eines Deutschen Forschungsnetzes e.V.*, welches die Fachhochschule *HAW Hamburg* und das *Max-Planck-Institut für Informatik* verwenden. Forschungseinrichtungen, welche nicht auf ein solches Forschungsnetz zurückgreifen, sind *Ruhr-Universitaet Bochum*, *RWTH Aachen*, *National Institute of Technology Karnataka* und *Stanford University*. Anhand der gescannten Ports lässt sich feststellen, dass ein großes Forschungsinteresse an dem DNS-Protokoll (53) besteht. Weiter sind bekannte Protokolle wie QUIC (443) und NTP (123) vertreten.

Forschungseinrichtung	Land	RIR Name	DNS Eintrag	Ports
Ruhr-Universitaet Bochum	DE	Ruhr-Universitaet Bochum	ports-measurements.softsec.ruhr-uni-bochum.de	5683, 5684, 53
HAW Hamburg	DE	Verein zur Foerderung eines Deutschen Forschungsnetzes e.V.	research-scan1.inet.haw-hamburg.de	53
Max-Planck-Institut für Informatik	DE	Verein zur Foerderung eines Deutschen Forschungsnetzes e.V.	inet-research-scan-1.mpi-inf.mpg.de	123
RWTH Aachen	DE	RWTH Aachen University	researchscan27.comsys.rwth-aachen.de researchscan36.comsys.rwth-aachen.de	53, 443
Universität Twente	NL	SURF B.V.	192-87-173-76.measurements.dacs.utwente.nl reactive-measurements.dacs.utwente.nl please.visit.www.openintel.nl	53
University of Southern Denmark	DK	Forskningsnettet - Danish network for Research and Education	srv-d1.sdu.dk	53
Technischen Universität Dänemark	DK	Forskningsnettet - Danish network for Research and Education	dtuscanner.compute.dtu.dk	7400, 5683
University of Cambridge	GB	Jisc Services Limited	cccc-scanner.cl.cam.ac.uk email-cccc-infra--cccc-scanner.cst.cam.ac.uk	1900, 123, 17, 389
Université Grenoble Alpes	FR	Renater	aix.u-ga.fr	53
Technische Hochschule Georgia	US	Georgia Institute of Technology		53
Tigard-Tualatin School District	US	Multnomah Education Service District	autohost66-154-208-12.ttsd.k12.or.us	500
National Institute of Technology Karnataka	IN	NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA	nitk-gw2-up-3.nitk.ac.in	8853, 784, 853
	CN	China Education and Research Network Center		53
Stanford University	US	Stanford University	research.esrg.stanford.edu	53

Abbildung 5.14: Übersicht der zu Forschungszwecken betriebenen Scans.

5.3 DoS/DDoS-Angriffe

Im nachstehenden Unterkapitel folgt eine genauere Untersuchung der im Cloud-Teleskop identifizierten (backscatter) DoS/DDoS-Angriffe.

Diese Untersuchung beginnt mit der Betrachtung der bei diesen Angriffen attackierten Ports. Hierzu ist in 5.15 die Anzahl an Angriffen visualisiert, welche den jeweilige Ports verwendet haben. Zum einfacheren Verständnis sind die Ports hierbei zu den entsprechenden Protokollen aufgelöst. Zu einem Angriff zusammengefasst werden alle Backscatter-Antworten, bei denen dieselbe Ziel-IP auf dem gleichem Ziel-Port angegriffen wurde. Zur Erstellung des Diagramms wurde eine Gruppierung der Angriffe anhand der Ziel-Ports vorgenommen.

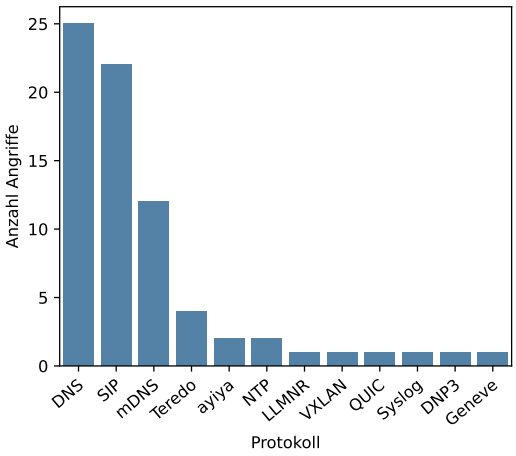


Abbildung 5.15: Verteilung der Angriffe anhand der verwendeten Portokolle.

Bei Betrachtung der Abbildung ist schnell festzustellen, dass die drei Protokolle *DNS*, *Session Initiation Protocol* und *Multicast DNS* zusammengefasst für über 85% aller registrierten Angriffe verwendet wurden. Wie in Abschnitt 3.5 beschrieben, eignen sich die drei genannten Protokolle zur Durchführung von *UDP-Flooding* Angriffen. Wie dort weiter beschrieben lassen sich die Protokolle *DNS* und *NTP* zudem mit Amplification-Attacken in Verbindung bringen. Gleiches gilt für *Multicast-DNS*. Hier kann durch die in Unterabschnitt 3.1.1 beschriebene ANY-Anfrage eine Verstärkung der Antwort erzielt werden. Da mDNS standardmäßig nur in lokalen Netzwerken verwendet wird, ist hier die Ausnutzung eines fehlerhaft konfigurierten mDNS-Responder notwendig.

Um weiterführende Informationen über die DoS/DDoS-Angriffe zu erhalten, folgt mit 5.16 und 5.17 eine Betrachtung unter der Einbeziehung des geografischen Aspekts. In 5.16 wird zunächst ein allgemeiner Überblick der betroffenen Länder dargestellt. Um die attackierten src-IPs geografisch einzuordnen, wurde auf die *country*-Spalte der *standard-location* Geo-Datenbank zurückgegriffen.

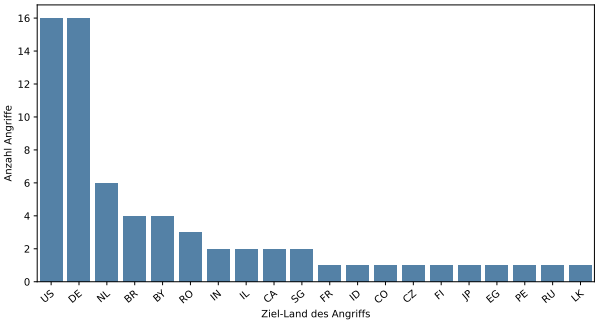


Abbildung 5.16: Verteilung der Angriffe auf die Top 20 am stärksten attackierten Länder.

Wie bereits oben bei den attackierten Ports beobachtet, konzentriert sich ein Großteil der Angriffe auf nur wenige Ziele. Die beiden Länder, Deutschland und USA, teilen sich hierbei die Spitzenposition unter den Staaten, auf die die meisten Angriffe verübt wurden. Durch das Zusammenfassen der ersten drei Länder ist zu erkennen, dass diese mit 56,7% mehr als die Hälfte aller Angriffe zu verzeichnen haben.

Um nach der Betrachtung der attackierten Länder nun einen geografischen Zusammenhang zu den Angreifern herzustellen, wurde die Heatmap 5.17 entwickelt. Wie in 4.5 bereits erläutert, sind in einem Internet-Teleskop nur solche Angriffe erkennbar, bei welchen der Angreifer die src-IP der Anfragen mittels IP-Spoofing fälscht. Wird diese src-IP von einer VM des Cloud-Teleskops verwendet, so sind die Paketantworten des angegriffenen Servers im Teleskop vorzufinden. In den Daten des Cloud-Teleskops sind also nur solche Backscatter-Antworten vorzufinden, bei welchen die gefälschte src-IP aus dem IP-Adressbereich von Digital-Ocean stammt. Unter Betrachtung der geografischen Komponente bedeutet dies, dass sich die src-IPs anhand der Serverstandorte von Digital-Ocean zusammenfassen lassen. Die Heatmap zeigt somit einen Vergleich zwischen dem angegriffenen Land und dem Serverstandort, dessen src-IP für den Angriff verwendet wurde. Die x-Achse steht hierbei für die Serverstandorte und die y-Achse für die attackierten Länder.

Zur Berechnung der Ausgangsdaten des Diagramms wurde ein entsprechender Python-Algorithmus entworfen. Als Eingabeparameter werden dem Algorithmus ein Datensatz und zwei zugehörige Vergleichsspalten übergeben. Als Datensatz wurde die in Abschnitt 4.5 beschriebene *backscatter.log* Datei verwendet. Zur Erstellung der Vergleichsspalten wurden die Quell- und Ziel-IPs dieses Datensatzes mit entsprechenden Geo-Informationen aus der *standard-location* Datenbank erweitert.

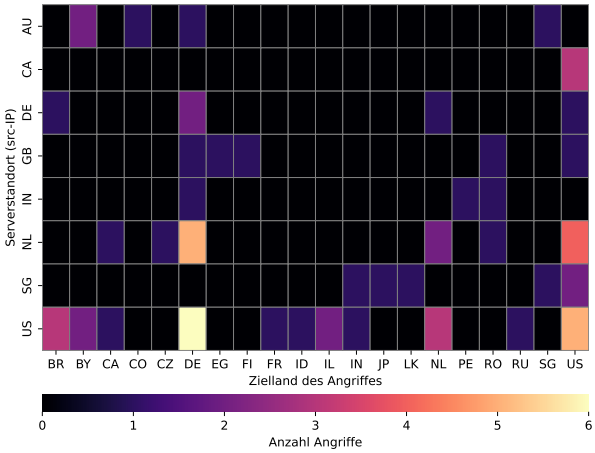


Abbildung 5.17: Anzahl gefundener Angriffe in Abhängigkeit des attackierten Landes sowie des geografischen Ursprungs der für den Angriff verwendeten src-IP.

Die Funktionsweise des Algorithmus besteht darin, die Zeilen von *backscatter.log* iterativ zu durchlaufen und jede Kombination der beiden Vergleichsspalten im Rückgabewert festzuhalten. Alle Kombinationen, für welche kein Treffer gefunden wird, werden entsprechend mit einem Nullwert gefüllt.

Bei Betrachtung der Heatmap ist zunächst festzustellen, dass gefälschte src-IPs von 8 verschiedenen Serverstandorten verwendet wurden. Hierbei weiter zu beobachten ist, dass die Serverstandorte Australien und Großbritannien

in dieser Menge enthalten sind, aber auf diese Länder selbst keine Angriffe verübt wurden. Weiter zu erkennen ist, dass die Serverstandorte in der USA die meisten Angriffe zu verantworten haben. Eine Besonderheit lässt sich bei dem Vergleich der Serverstandorte von den USA und Deutschland feststellen. Wie zu 5.16 beschrieben, wurden gleich viele Angriffe auf beide Länder verübt. Mit Blick auf die Heatmap ist allerdings festzustellen, dass von den IP-Adressen der Server in der USA deutlich häufiger Angriffe durchgeführt wurden. Ein deutlicher regionaler Zusammenhang zwischen dem Serverstandort und der Anzahl registrierter Angriffe auf dasselbe Land lässt sich nicht feststellen. So wurden beispielsweise src-IPs von Servern in der Niederlande deutlich häufiger für Angriffe auf Deutschland oder die USA verwendet als auf die Niederlande selbst. Ähnliches gilt für Kanada. Hier sind es Ziele in den USA, welche am häufigsten von den Angriffen betroffen sind.

Um nun einen tieferen Einblick darin zu erhalten, welche Organisationen/Unternehmen angegriffen wurden, sind in 5.18 die Top 20 RIR-Namen aufgeführt, von denen die höchste Anzahl an Backscatter-Antworten registriert wurden. Zur Berechnung der Ausgangsdaten wurden die Backscatter-Antworten aus *backscatter.log* mit den zugehörigen RIR-Namen erweitert. Diese Zuordnung wurde anhand der gefälschten src-IPs durchgeführt. Um weiterführende Informationen über die Opfer der Angriffe zu erhalten, wurden für die entsprechenden src-IPs DNS-Reverse-Anfragen durchgeführt. Konnte hierbei ein Treffer gefunden werden, so wurde dieser als x-Wert im Diagramm verwendet. Konnte hingegen kein Treffer gefunden werden, so wurde der RIR-Name verwendet.

Bei Betrachtung der DNS-Reverse-Einträge ist auffällig, dass diese oft die Struktur automatisch generierter Cloud-Maschinen besitzen. Die Subdomain steht hierbei jeweils für einen Hosting-Anbieter. Der Hostname ist entweder eine zufällige Zahlenkombination oder die IP der gemieteten Maschine selbst. Um mehr Informationen über die einzelnen Maschinen und deren Verwendung zu erfahren, wurden die Ports mit dem Kommandozeilentool *nmap*[53] untersucht. Einige der IPs konnten allerdings nicht erreicht werden, oder anhand der Ausgaben konnten keine sinnvollen Schlüsse gezogen werden.

Zu *vmi1278843.contaboserver.net.* konnte ein offener *ssh*-Port (22) und ein offener *DNS*-Port (53) gefunden werden. Hingegen konnten zu *v281302.hosted-by-vdsina.com.* offene Ports für *DNS* (53), *http* (80) und *https* (443) gefunden werden, sowie die gefilterten Ports 135-139. Zu der Cloud-Maschine *216-10-250-164.webhostbox.net.* konnte ein offener *SIP* Port (5060) gefunden werden. Bei Betrachtung der im Diagramm aufgeführten RIR-Namen sind neben vielen kleineren Anbietern auch große Unternehmen wie Microsoft oder Google vertreten. Eine Vielzahl der sonst aufgeführten Cloud-Anbieter haben ihren Firmensitz in den USA. Ein in Deutschland ansässiges Unternehmen ist hingegen *aurologic GmbH*, welches Produkte aus dem Bereich Cloud-Hosting anbietet.

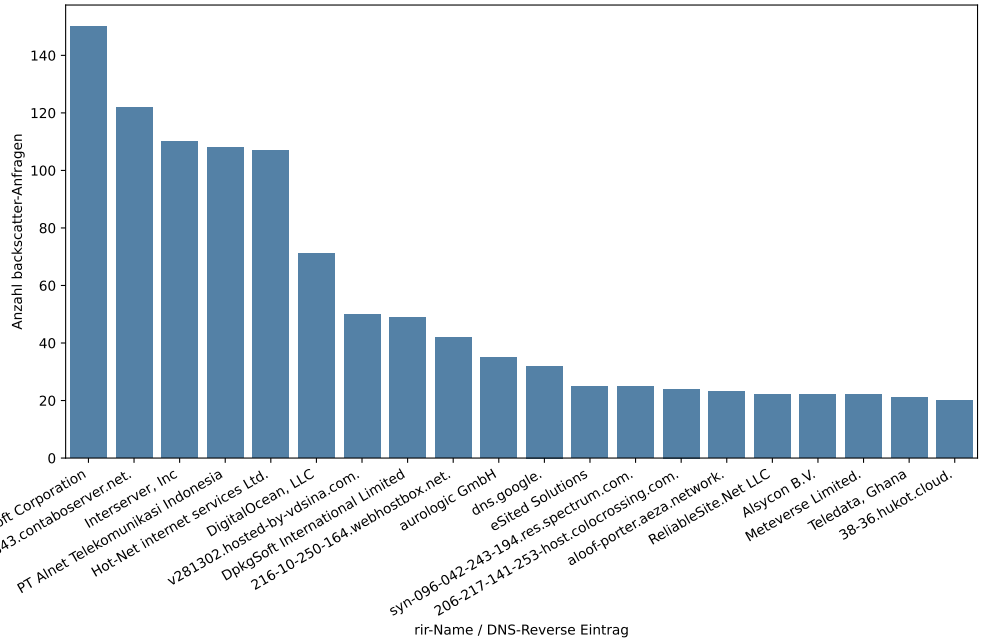


Abbildung 5.18: Verteilung der registrierten Backscatter-Antworten auf die Top 20 am stärksten attackierten RIR-Namen.

Abschließend zu diesem Unterkapitel wurde eine Analyse durchgeführt, welche das Ziel hat, den Angriffsverlauf genauer zu verstehen. Hierzu wurde für alle zuvor gefundenen Angriffe das in Abschnitt 3.2 beschriebene ICMP-Szenario überprüft. Hierbei wird ausgenutzt, dass ein Server standardmäßig mit einer ICMP *destination unreachable* (Type 3) Nachricht antwortet, sobald dieser eingehenden TCP-Anfragen (aufgrund einer temporären Überlastung) nicht beantworten kann. Werden für den Angriff nun sowohl TCP- als auch UDP-Pakete verwendet und wird hierbei auf die gleichen IP-Spoofing-Adressen zurückgegriffen, so sind die ICMP-Nachrichten in den Teleskop-Daten wiederzufinden. Diese Nachrichten werden dann so lange von Cloud-Teleskop aufgezeichnet, bis der Angriff nachlässt und der Server eingehende Anfragen wieder bearbeiten kann. Der Zeitraum des Serverausfalls lässt sich somit anhand der Zeitpunkte des zuerst und zuletzt registrierten *destination-unreachable*-Paketes bestimmen. Da die verwendeten Maschinen des Cloud-Teleskops nur für eine begrenzte Zeit aktiv sind, muss zudem sichergestellt werden, dass das Ende der empfangenden ICMP-Nachrichten nicht auf das Mietende der Cloud-Maschine zurückzuführen ist. Um dies auszuschließen, wurden die in Abschnitt 3.4 beschriebenen *descriptor*-Dateien verwendet. Diese werden vom Cloud-Teleskop angelegt und enthalten Metadaten der verwendeten Cloud-Maschinen. Für diese Analyse wurde der *deletion*-Zeitstempel dieser Dateien verwendet. Die empfangenen ICMP-Nachrichten des Cloud-Teleskops wurden aus der conn.log gefiltert. Die Funktionsweise des Algorithmus besteht nun darin, die einzelnen Angriffe iterativ zu durchlaufen und anhand der gefälschten src-IP alle zugehörigen ICMP-Nachrichten zu ermitteln. Diese ICMP-Einträge

werden anschließend nach dem *destination-unreachable* Typ gefiltert. Nach einer Überprüfung mit dem *deletion*-Zeitstempel erfolgt die Berechnung des Zeitraums der ICMP-Nachrichten.

Nach dem Durchlaufen aller gefundenen Angriffe, konnten die Bedingungen des beschriebenen ICMP-Szenarios lediglich für einen Angriff nachgewiesen werden. Dieser Angriff erfolgte auf die IP *102.70.66.6*, welche zum RIR-Namen *TELEKOM NETWORKS MALAWI LTD* gehört. In dem entsprechenden Zeitraum konnten vier ICMP Type-3 Pakete registriert werden. Anhand dieser Pakete konnte eine Zeitspanne von 20 Minuten und 21 Sekunden ermittelt werden. *TELEKOM NETWORKS MALAWI LTD* ist ein Telekommunikationsanbieter in Malawi (Ostafrika), welcher unter anderem für das dortige Mobilfunknetz verantwortlich ist. Eine Untersuchung der UDP-Ports mittels des Tools *nmap* hat offene Ports für DNS (53) und NTP (123) aufgedeckt.

5.4 Vergleich CAIDA-Teleskop

Dieses Kapitel befasst sich mit einem Vergleich zwischen dem verwendeten Cloud-Teleskop und dem CAIDA-Teleskop. Hierzu wird überprüft, welche der zuvor im Cloud-Teleskop gefundenen Phänomene in den Daten des CAIDA-Teleskops nachzuweisen sind. Um einen validen Vergleich durchführen zu können, werden für diese Analyse nur Daten des klassischen Teleskops verwendet, welche in demselben Zeitraum aufgezeichnet wurden. Wichtig zu beachten ist hierbei, dass zwischen den Zeitpunkten am 25.01.2025 um 11 Uhr und am 27.01.2025 um 19 Uhr keine Daten in den entsprechenden Swift-Containern der CAIDA-Infrastruktur vorzufinden waren. Dies lässt sich vermutlich auf einen temporären Ausfall des Teleskops zurückführen. Aus diesem Grund wurden für alle folgenden Vergleiche die Daten des Cloud-Teleskops so angepasst, dass keine Verbindungen dieses Zeitraums einbezogen werden. Zur Erkennung von Scans sowie von Backscatter-Traffic wurden für das CAIDA-Teleskop auf die gleichen Verfahren zurückgegriffen, welche zuvor für das Cloud-Teleskop verwendet wurden.

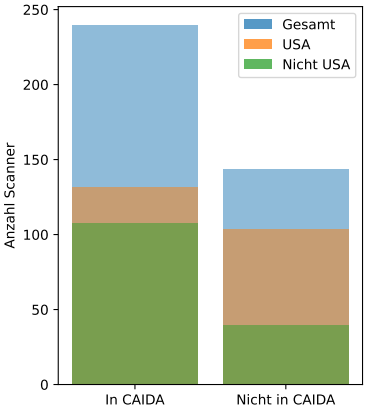
Der Vergleich der beiden Teleskope ist in zwei Unterkapitel gegliedert. Hierzu folgt zunächst eine Gegenüberstellung der gefundenen Scans. Daran anschließend werden der Backscatter-Traffic und die so identifizierten Angriffe genauer analysiert.

5.4.1 Vergleich Scanning

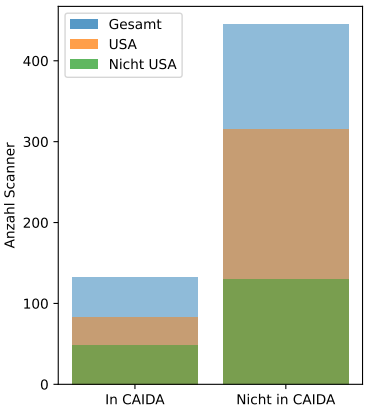
Um die für diesen Vergleich notwendigen Daten des CAIDA-Teleskops zu laden, wurden die in Abschnitt 4.3 beschriebenen Shell-Scripte verwendet. Diese Scripte wurden auf der CAIDA-VM jeweils für Port- und Address-Scans durchgelaufen. Zum Vorfiltern der zu ladenden Pcap-Dateien wurde den Scripten jeweils eine Liste an IP-Adressen mitgegeben. Diese IP-Adressen entsprechen

hierbei den src-IPs, welche für Port- und Address-Scans in dem Cloud-Teleskop gefunden wurden. Die von Zeek im jedem Durchlauf erzeugte conn.log Datei wurde während des Iterationsschrittes so angepasst, dass diese nur aus den folgenden Spalten bestehen: *ts*, *id.orig_h*, *id.orig_p*, *id.resp_h*, *id.resp_p*, *proto* und *conn_state*. Die Laufzeit der Scripte betrug jeweils ca. 2,5 Tage. Wie eingangs beschrieben, entsprechen alle weiteren Verfahren denen, welche zuvor für die Identifizierung von Scans beschrieben wurden.

Somit sind auch für die Daten des CAIDA-Teleskops die beiden Parameter des zeitlichen Intervalls sowie des Anfragen-Grenzwertes festzulegen. Wie in Abschnitt 4.4 beschrieben, liegt der Anfragen-Grenzwert der Address-Scans bei 100 und der für Port-Scans bei 25. Das Intervall ist hierbei für beide Scan-Typen auf 8 Minuten festgelegt. In 5.19a sind für Address- und Port-Scanner einige Vergleiche beider Teleskope visualisiert. Hierbei steht jeweils die Summe der beiden Säulen, welche mit dem Label *Gesamt* gekennzeichnet sind, für die Anzahl aller im Cloud-Teleskop gefundenen Scanner. Dieser Wert wird in den Diagrammen in zwei Säulen aufgeteilt. Die Säule mit dem x-Wert *In CAIDA* steht hierbei für die Anzahl an Scannern, welche im CAIDA-Teleskop gefunden wurden. Die Säule mit dem x-Wert *Nicht in CAIDA* beschreibt hingegen die Anzahl der Scanner, welche nicht im CAIDA-Teleskop vorzufinden waren. Für jede der beiden Gruppen wurde anschließend eine geografische Betrachtung durchgeführt. Hierzu wurde für alle src-IPs der zugehörige Kontinent ermittelt. Diese Zuordnungen wurden auf Grundlage der *standard-location* Geo-Datenbank durchgeführt. Anhand dieses Wertes ist jede Säule in zwei Anteile aufgeteilt. Dies ist der Anteil an Scannern, deren Ursprung in den USA liegt und der Anteil, deren Ursprung außerhalb der USA liegt.



(a) Address-Scanner



(b) Port-Scanner

Abbildung 5.19: Vergleich erkannter Scanner beider Teleskope sowie deren geografische Einordnung.

Bei Betrachtung der Address-Scans in 5.19a ist zunächst festzustellen, dass mit 244 Scannern (62%) deutlich mehr als die Hälfte aller Scanner im CAIDA-Teleskop gefunden wurden. Bei der geografischen Betrachtung lässt sich kein Zusammenhang zwischen den im CAIDA-Teleskop gefundenen Scannern und dem Ursprung in den USA feststellen. So sind hier die Anteile der beiden geografischen Gruppen nahezu ausgewogen. Bei den Scannern, die ausschließlich im Cloud-Teleskop erfasst wurden, dominiert der Anteil derjenigen mit

Ursprung in den USA.

Bei der Analyse der Port-Scans in 5.19b lassen sich hingegen andere Verteilungen feststellen. So wurde hier mit 22% nur jeder fünfte Scanner in den Daten des CAIDA-Teleskops gefunden. Ein geografischer Zusammenhang mit dem Ursprung der Scanner lässt sich allerdings auch hier nicht feststellen. So überwiegt bei beiden Gruppen der Anteil an Scannern, deren Ursprung in der USA liegt.

Die beiden Tabellen aus 5.20 bieten abschließend einen Überblick darüber, welche Scans in den Daten des CAIDA-Teleskops gefunden wurden und welche nur im Cloud-Teleskop nachweisbar waren. Die Tabelle 5.20a steht hierbei für die Address-Scans und in 5.20b sind die Port-Scans aufgelistet. In den beiden Spalten der Tabellen sind jeweils die RIR-Namen der Top 10 Scanner beider Teleskope aufgeführt. Um diese beiden Gruppen zu erhalten, wurde zunächst von den im Cloud-Teleskop identifizierten Scannern die Teilmenge entfernt, welche auch im CAIDA-Teleskop gefunden wurden. Eine Sortierung der übrig gebliebenen Scanner anhand der Anzahl der zugehörigen Scans ergibt die Spalte *Nicht in CAIDA gefunden*. Die Spalte *In CAIDA gefunden* ergibt sich durch die gleiche Sortierung der im CAIDA-Teleskop gefundenen Scanner. Zu beachten ist, dass gleiche RIR-Namen in beiden Spalten der Tabellen vorkommen. Dies lässt sich damit begründen, dass die src-IPs verschiedener Scanner zu demselben RIR-Namen gehören können. Ein solcher Fall zeigt sich beispielsweise bei dem RIR-Namen *DigitalOcean, LLC*.

In CAIDA gefunden	Nicht in CAIDA gefunden
Arbor Networks, Inc.	Perfecto Consultoria E Apoio Administrativo
Optibounce, LLC	ssd networks limited
HLL LLC	HKN GmbH
Pfcloud UG	DigitalOcean, LLC
DediOutlet, LLC	WholeSale Internet, Inc.
FranTech Solutions	DK Microsoft Corporation
ColocationX Ltd.	China Mobile Communications Group Ltd.
DigitalOcean, LLC	G Arbor Networks, Inc.B
SingleHop LLC	F ReliableSite.Net LLCR
EKABI	IP Volume inc

(a) Address-Scans

In CAIDA gefunden	Nicht in CAIDA gefunden
AntiDDoS Solutions LLC	Mobile Communication Company of Iran PLC
DigitalOcean, LLC	Information Technology Company (ITC)
NSEC - Sistemas Informaticos, S.A.	DigitalOcean, LLC
Cogent Communications	Rethem Hosting LLC
Zenlayer Inc	CHINANET-BACKBONE
Rethem Hosting LLC	ssd networks limited
JSC Selectel	Aria Shatel PJSC
xTom	PARDIS FANVARI PARTAK LTD
SRMAK TECHNOLOGICAL SYSTEM LIMITED	F ReliableSite.Net LLCR
SpectralP B.V.	IP Volume inc

(b) Port-Scans

Abbildung 5.20: Übersicht der Top 10 RIR-Namen beider Teleskope, welche die meisten Scans zu verantworten haben.

5.4.2 Vergleich DoS/DDoS-Angriffe

Nachstehend wird verglichen, welche der im Cloud-Teleskop gefundenen Angriffe auch in den Daten des CAIDA-Teleskops nachzuweisen sind. Um den vom CAIDA-Teleskop erfassten Backscatter-Traffic aus der CAIDA-Infrastruktur zu laden, wurden die im vorherigen Unterkapitel beschriebenen Schritte in ähnlicher Weise durchgeführt. Zum Vorfiltern der Daten wurden die Ziel-IPs verwendet, zu welchen ein Angriff im Cloud-Teleskop nachgewiesen wurde. Zur Erkennung der Backscatter-Antworten wurde das in 4.5 beschriebene Verfah-

ren angewandt. Da die verwendete CAIDA-VM nur über einen begrenzten Speicherplatz verfügt, wurde das Shell-Script zudem so angepasst, dass die Backscatter-Erkennung in jedem Iterationsschritt durchgeführt wird. So werden in jedem Durchlauf nur die Verbindungen der von Zeek erzeugten conn.log gespeichert, welche als Backscatter-Traffic klassifiziert wurden.

Im Diagramm 5.21 erfolgt ein Vergleich der beiden Teleskope anhand der Anzahl erkannter Angriffe sowie der Anzahl erfasster Backscatter-Antworten. Hierzu visualisiert die erste Säule die Angriffe, welche in beiden Teleskopen gefunden wurden und die Teilmenge an Angriffen, welche nur im Cloud-Teleskop erfasst wurde. Auch die zweite Säule lässt sich in zwei Bereiche einteilen. Der *Cloud-Teleskop* Anteil steht hierbei für die Anzahl der im Cloud-Teleskop erfassten Backscatter-Antworten. Der *CAIDA-Teleskop* Anteil hingegen steht für die Anzahl der Backscatter-Antworten des Cloud-Teleskops, zu welchen die jeweilige src-IP in den Daten des CAIDA-Teleskops gefunden wurde. Bei dem Diagramm ist die unterschiedliche Skalierung der beiden Säulen zu beachten .

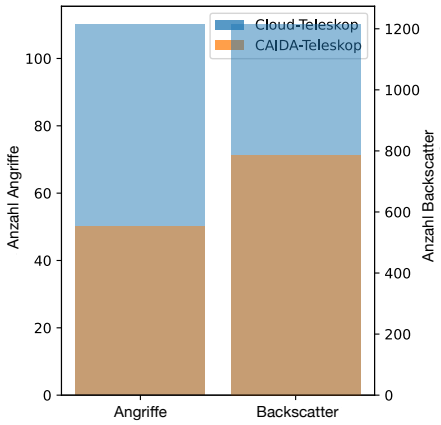


Abbildung 5.21: Vergleich beider Teleskope anhand der Anzahl gefundener Angriffe sowie der Anzahl registrierter Backscatter-Antworten.

Bei der Betrachtung der erkannten Angriffe lässt sich erkennen, dass nur weniger als die Hälfte aller Angriffe (45%) im CAIDA-Teleskop nachzuweisen sind. Hinsichtlich der Backscatter-Anfragen ist jedoch festzustellen, dass die src-IPs von über 64% aller Backscatter-Antworten in den Daten des CAIDA-Teleskops gefunden wurden. Aus der Kombination beider Beobachtungen lässt sich nun schließen, dass die geringe Anzahl der im CAIDA-Teleskop gefundenen Angriffe einen erheblichen Anteil der Backscatter-Antworten des Cloud-Teleskops zu verantworten hat.

Abschließend gibt die Abbildung 5.22 einen genaueren Überblick darüber, welche angegriffenen Ziele in den Daten des CAIDA-Teleskops gefunden wurden. Hierzu sind die Top 20 RIR-Namen / DNS-Einträge aufgeführt, zu denen die größte Anzahl an Backscatter-Antworten registriert wurde. Die Ermittlung der Ausgangsdaten erfolgte hierbei mit denselben Verfahren, welche zur Erstellung von 5.18 vorgestellt wurden.

Bei der Betrachtung des Diagramms ist die hohe Anzahl erfasster Backscatter-Anfragen im CAIDA-Teleskop zu erkennen. Da sich diese in der Größenordnung von 10^7 bewegen, ist die Skalierung der y-Achse entsprechend angepasst. Ein Vergleich der aufgeführten RIR-Namen / DNS-Einträge mit denen aus 5.18

zeigt sowohl Übereinstimmungen als auch zahlreiche Unterschiede. So sind neun der 20 Einträge in beiden Diagrammen zu finden. Zu diesen Einträgen gehören unter anderem *eSited Solutions*, *vmi1278843.contaboserver.net.*, *PT Alnet Telekomunikasi Indonesia* oder *DigitalOcean, LLC*. Weiter lassen sich unterschiedliche Positionierungen der Einträge in den beiden Diagrammen feststellen. So steht *eSited Solutions* mit über 25 Millionen Anfragen an erster Stelle aller im CAIDA Teleskop gefundenen Angriffe. In 5.18 hingegen ist dieser RIR-Name im Mittelfeld vorzufinden.

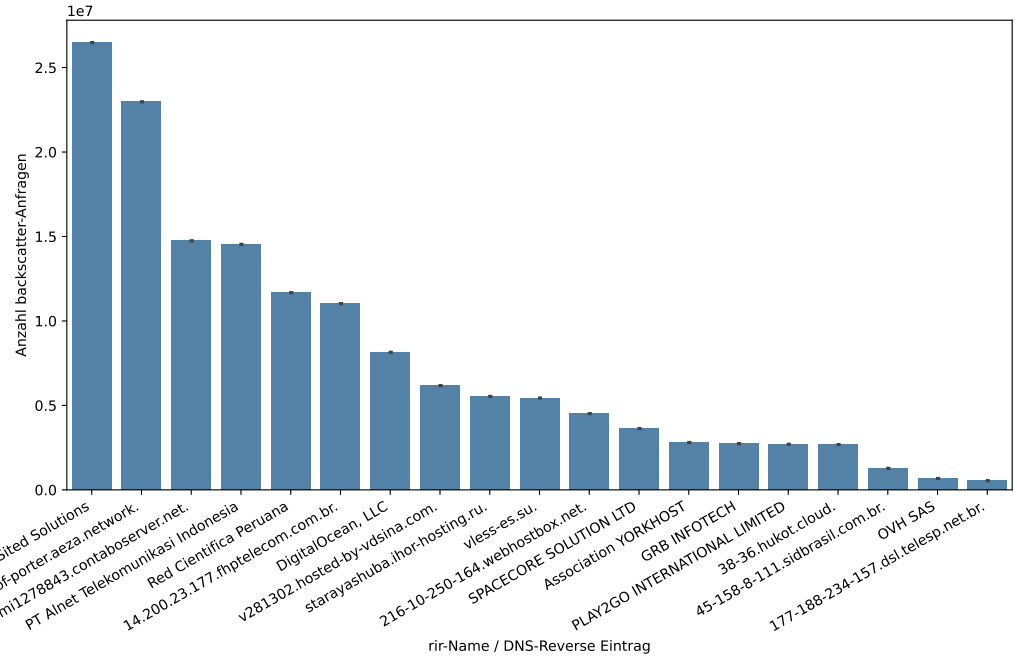


Abbildung 5.22: Verteilung der im CAIDA-Teleskop registrierten Backscatter-Antworten auf die Top 20 am stärksten attackierten RIR-Namen.

6 Diskussion

Im Folgenden findet eine Diskussion über die im vorherigen Kapitel erläuterten Ergebnisse statt. Die Unterkapitel der Diskussion orientierten sich an den in Kapitel 1 vorgestellten Forschungsfragen. Somit beginnt dieses Kapitel mit der Begründung der zur Erkennung von Scans festgelegten Parameter. Daran anschließend folgt eine Diskussion über die im Cloud-Teleskop gefundenen Phänomene. Abschließend werden die Ergebnisse des Vergleichs beider Teleskope genauer interpretiert.

6.1 Herleitung der Parameter zur Identifizierung von Scans

Zur Herleitung eines zeitlichen Intervalls sowie eines minimalen Grenzwertes an Anfragen werden zunächst die Ergebnisse aus Unterabschnitt 5.2.1 genauer interpretiert. In diesem Unterabschnitt sind die Ergebnisse der drei Heatmaps 5.3, 5.4 und 5.5 dargestellt.

Zunächst konnte zu 5.3 festgestellt werden, dass sich die Anzahl gefundener Scans bei einem steigenden Anfragen-Grenzwert verringert. Dies lässt sich damit begründen, dass jeder Scanner über eine bestimmte Scan-Rate verfügt. Dies bedeutet, dass ein Scanner pro Zeiteinheit nur eine bestimmte Anzahl an Hosts/Ports scannt[41]. Wird der Anfragen-Grenzwert höher gesetzt als die jeweilige Scan-Rate, so wird dieser Scanner nicht mehr als solcher erkannt. Andersherum führt ein niedriger Anfragen-Grenzwert dazu, dass auch Anfragen von Fehlkonfigurationen fälschlicherweise als Scan erkannt werden.

Zunächst im Widerspruch zu den Begründungen einer solchen Scan-Rate, steht das zu 5.3a erkannte Phänomen. Demnach steigt die Anzahl gefundener Scans nur bis zu einer bestimmten Intervallgrenze. Nach Überschreiten dieser Grenze sinkt die Anzahl gefundener Scans. Dieses Verhalten lässt sich damit begründen, dass ab dieser Intervallgrenze Scans mit derselben src-IP zu einem einzigen Scan zusammengefasst werden, während sie bei einem kürzeren Intervall als separate Scans erkannt wurden. Somit ergibt sich die Differenz der Werte zweier aufeinanderfolgender Intervalle aus den in diesem Zeitabschnitt neu identifizierten Scans und den ab diesem Zeitabschnitt zusammengefassten Scans.

Mit der Anzahl neu gefundener Scanner befassen sich die Ergebnisse zu den Heatmaps aus 5.4. Die Erkenntnis, dass die Anzahl neu gefundener Scanner bei steigendem Anfragen-Grenzwert abnimmt, ist mit der gleichen, zuvor be-

schriebenen, Begründungen erklärbar. Deutlich interessanter ist hier der Verlauf der Werte bei steigendem Intervall. So zeigt dieser Verlauf, wie viele Scans ab dem jeweiligen Intervall erstmalig den Anfragen-Grenzwert überschreiten, also genug Anfragen versenden, um als Scan erkannt zu werden. Der zu den Port-Scannern in 5.4b beschriebene Verlauf für den Anfragen-Grenzwert von 4 lässt sich damit begründen, dass dieser Wert so klein ist, dass nahezu alle Scanner/Fehlkonfigurationen diesen schon bei dem niedrigsten Intervall von einer Minute erreichen.

Der Anfragen-Grenzwert soll also gerade so groß gesetzt werden, dass für die weitere Analyse möglichst viele Scans betrachtet werden, ohne hierbei eine große Anzahl an Fehlkonfigurationen miteinzubeziehen. Um nun für Address- und Port-Scans einen passenden Anfragen-Grenzwert festzulegen, kann eine Sensitivitätsanalyse der beschriebenen Heatmaps herangezogen werden. Bei Betrachtung der Port-Scans in 5.3b ist zu erkennen, dass bei einem Anfragen-Grenzwert von 4 eine überdurchschnittlich große Dichte im Vergleich zu allen anderen Werten vorliegt. So sind für diesen Anfragen-Grenzwert dauerhafte Werte über 5000 zu verzeichnen. Wie zuvor beschrieben, entspricht der Anfragen-Grenzwert von vier auch bei Betrachtung der neu erkannten Scanner in 5.4b nicht dem sonst vorliegenden Muster. Somit folgt aus der Sensitivitätsanalyse, dass für die Port-Scans der Anfragen-Grenzwert auf den nächstgrößten Wert von 6 festgelegt wurde.

Für die Address-Scans lassen sich durch eine solche Analyse keine erheblichen Ausreißer oder Abweichungen feststellen. Deshalb können gerade für diesen Scan-Typ mehrere Anfragen-Grenzwerte in Betracht gezogen werden. Eine Auswahl kann somit anhand der Analysen getroffen werden, welche mit den erkannten Scans durchgeführt werden sollen. Da für viele der in dieser Arbeit verwendeten Analysen vergleichbare Werte der beiden Scan-Typen vorteilhaft sind, wurde für Address-Scans derselbe Anfragen-Grenzwert festgelegt, wie bei den Port-Scans.

Um nun für diesen Anfragen-Grenzwert ein entsprechendes Intervall herzuleiten, hilft eine Interpretation der zu 5.6 dargestellten Ergebnisse. Die Koordinatensysteme visualisieren hierbei den Verlauf der normalisierten Funktionswerte der zuvor diskutierten Heatmaps für den Anfragen-Grenzwert von 6. Zunächst für beide Koordinatensysteme festzuhalten ist, dass die Scandauer eine monotone und gleichmäßige Steigung aufweist. So sind anhand dieses Verlaufes keine Intervallwerte im vorhinein auszuschließen.

Bei Betrachtung Funktionswerte der neu gefundenen Address-Scanner ist zwischen den Intervallwerten von 7 und 8 ein signifikanter Rückgang festzustellen. Somit werden ab diesem Intervallwert nur noch deutlich weniger neue Scanner in den Daten gefunden. Für den Funktionsverlauf der identifizierten Scans ist hingegen zu beobachten, dass eine deutliche Dekrementierung zwischen den Intervallwerten 4 und 5 zu verzeichnen ist. Dies hat zu bedeuten, dass ab diesem Intervall eine Vielzahl der zuvor einzeln betrachteten Scans

anhand der src-IP zusammengefasst werden. Aus der Kombination beider Betrachtungen können somit die Werte von 5 bis 7 für ein entsprechendes Intervall herangezogen werden. Für diese Arbeit wurde das Intervall ohne weitere Begründungen auf die Mitte dieser drei Werte festgelegt.

Für die normalisierten Funktionsverläufe der Port-Scans in 5.6b kann eine ähnliche Interpretation durchgeführt werden. Hier bei dem Verlauf der neu erkannten Scanner nahezu konstante Funktionswerte zwischen den Intervallwerten von 3 und 7 festzustellen. Ab dem Wert von 7 ist eine Abnahme zu verzeichnen, welche bei dem Wert von 9 im Nullpunkt endet. Bei Betrachtung des Funktionsverlaufs der Anzahl erkannter Scans ist eine monotone Steigung festzustellen. Dies bedeutet, dass zu jedem neuen Intervallwert mehr Scans gefunden wurden, als zuvor gefundene Scans zusammengefasst werden. Somit werden anhand dieses Verlaufes keine Intervallwerte ausgeschlossen. Für das zeitliche Intervall wurde somit der Wert von sieben gewählt, da wie oben beschrieben, bis zu diesem Wert eine konstante Anzahl neu erkannten Scans vorliegt.

6.2 Bewertung der im Cloud-Teleskop gefunden Phänomene

Im Folgenden wird nun eine Auswertung der im Cloud-Teleskop gefundenen Phänomene präsentiert. Das Unterkapitel beginnt hierfür mit einer Interpretation der in den Scans festgestellten Merkmale. Anschließend folgt eine Auseinandersetzung mit den dargestellten Ergebnissen der Dos/DDoS-Angriffe.

6.2.1 Bewertung Scanning

Im Folgenden werden zunächst die zu 5.7 und 5.8 gefundenen Merkmale genauer betrachtet. Das Diagramm 5.7 visualisiert hierbei den Verlauf identifizierter Scans über den gesamten Analysezeitraum. Ähnlich zeigt 5.8 die Entwicklung der zu diesen Scans zugehörigen Anfragen.

Zunächst ist anhand beider Diagramme festzustellen, dass das Cloud-Teleskop über den gesamten Analysezeitraum ohne Ausfälle gearbeitet hat und kontinuierlich Pakete aufgezeichnet wurden. Bei der Betrachtung der identifizierten Scans in 5.7 lassen sich weder für Address- noch für Port-Scans wiederkehrende Muster feststellen. Ebenso wenig erkennbar ist ein unterschiedliches Verhalten an Wochentagen im Vergleich zum Wochenende. Da der Analysezeitraum von 14 Tagen sowohl zusammenhängende Wochentage als auch Wochenenden umfasst, kann davon ausgegangen werden, dass solche Muster auch bei Betrachtung eines größeren Zeitraums nicht zu erkennen sind. Eine ähnliche Beobachtung und Argumentation gilt für Funktionsverläufe der registrierten Anfragen in 5.8.

Im Anschluss folgt eine genauere Betrachtung der erkannten Ausreißer. Zunächst ist festzustellen, dass alle Ausreißer aus 5.7 auch in dem Verlauf der erfassten Anfragen wiederzufinden sind. Die Erklärung hierfür ist recht trivial. So bedingt eine hohe Anzahl an Scans aufgrund des festgelegten Anfragen-Grenzwertes zwangsläufig eine hohe Anzahl an Scan-Anfragen. Interessant ist allerdings, dass für diese Ausreißer die Gegenrichtung nicht gilt. Als Extrembeispiel dient hierfür der Ausreißer, den *gcommer.com* zu verantworten hat. Somit kann nachgewiesen werden, dass *gcommer.com* extrem aggressive Port-Scans durchführt, bei denen von wenigen Scannern extrem viele Anfragen ausgehen.

Eine weitere interessante Schlussfolgerung lässt sich zu den Hochpunkten am 29.01 und am 30.01 aufstellen. So ist erkennbar, dass zu den beiden Zeiträumen in 5.7 nur Hochpunkte der Port-Scans zu verzeichnen sind. In 5.8 hingegen sind für die gleichen Zeiträume für beide Scan-Typen Ausreißer festzustellen. Anhand der RIR-Namen ist zu erkennen, dass beide Ausreißer der Address-Scans auf dieselbe IP-Adresse zurückzuführen sind. Auffällig hierbei ist, dass diese IP aus dem Digital-Ocean IP-Adressbereich stammt. Die beiden Ausreißer der Port-Scan-Anfragen stammen nicht von dieser IP, sondern haben ihren Ursprung größtenteils im Iran. Weiter zu erkennen ist, dass die Ausreißer der beiden unterschiedlichen Scan-Typen nahezu zeitgleich beginnen und enden. Dadurch, dass alle oben genannten Auffälligkeiten nicht nur in einem Zeitraum, sondern gleich in zwei getrennten Zeiträumen nachzuweisen sind, kann die begründete Vermutung aufgestellt werden, dass sowohl die Ausreißer der Port- als auch die der Address-Scans auf denselben Betreiber zurückzuführen sind. Dies würde weiter bedeuten, dass die Port-Scans von Servern im Iran durchgeführt wurden, für die Address-Scans allerdings Cloud-Server von Digital-Ocean gemietet wurden.

Der folgende Abschnitt befasst sich mit den Ergebnissen, welche zur geografischen Betrachtung des Scan-Verhaltens in Unterabschnitt 5.2.3 dargestellt sind. Hierbei werden zunächst die Beobachtungen zum Scan-Verhalten der einzelnen Länder mit den Ergebnissen einer anderen Studie verglichen. So ist festzustellen, dass die zu 5.9 aufgestellten Beobachtungen nur teilweise mit der Studie von Durumeric *et al.*[8] übereinstimmen. Aus dem Diagramm geht hervor, dass China nur einen sehr geringen Anteil aller Scans zu verantworten hat. Dieser nur sehr geringe Anteil chinesischer Scans ist nicht in den Ergebnissen des Papers wiederzufinden. Eine Erklärung hierfür könnte sein, dass dieses Paper hauptsächlich TCP-Traffic betrachtet und das Scan-Verhalten der Länder bei UDP grundsätzlich ein anderes ist. So kann vermutet werden, dass von China ein verstärktes Interesse am TCP-Scanning besteht, während UDP weniger relevant ist. Um diese Vermutung zu validieren, wäre die Betrachtung eines weiteren Analysezeitraums. Die Beobachtung, dass aus dem Iran ausschließlich Port-Scans zu verzeichnen sind, lässt sich mit der im vorherigen Abschnitt aufgestellten These begründen. Demnach verwenden Organisationen des Iran gemietete Cloud-Maschinen zur Durchführung von Address-Scans.

Bei der Betrachtung der gescannten Protokolle in 5.10 lässt sich eine Begründung für viele dieser Protokolle finden. So lässt sich die enorme Anzahl gescannter DNS-Ports mit den in Abschnitt 3.5 beschriebenen *DNS-Amplification* Angriffen begründen. Hier besteht sowohl für Sicherheitsforscher als auch für potenzielle Angreifer ein großes Interesse daran, Scans durchzuführen. Eine ähnliche Argumentation gilt für die Vielzahl an Address-Scans, welche den NTP-Port priorisieren. Das *Constrained Application Protocol* (Port 5683) ist, wie in Abschnitt 3.6 beschrieben, ein Web-Transfer-Protokoll, welches speziell für Internet-of-Things-Geräte entwickelt wurde. Wie in 5.14 zu erkennen ist, lassen sich eine Vielzahl der Scans auf dieses Protokoll mit dem großen Forschungsinteresse der Ruhr-Universität-Bochum begründen. Somit kann vermutet werden, diese von dieser Universität aktuell Internetforschungen im Bereiche Internet-of-Things durchgeführt werden. Das *Session Initiation Protocol* wird, wie in Abschnitt 3.6 beschrieben, hauptsächlich für Voice-over-IP-Anwendungen verwendet. Das hohe Scanaufkommen dieses Protokolls lässt sich auch in den Forschungsergebnissen des Papers von Durumeric *et al.*[8] wiederfinden. Wie dort beschrieben, wird dieses Protokoll mit vielen Sicherheitslücken in Verbindung gebracht. Durch diese Sicherheitsrisiken kann somit das Scan-Interesse begründet werden.

In dem folgenden Abschnitt werden die Ergebnisse zu 5.11 genauer interpretiert. In diesem Diagramm wird ein Zusammenhang zwischen dem geografischen Ursprung der Address-Scans und den für diese Scans verwendeten Protokollen hergestellt. Dort zu erkennen ist, dass die Säulen, welche zu Deutschland gehören, bei fast allen Protokollen hohe Werte zeigen. Ein solches Verhalten wäre auch für die anderen Länder zu vermuten sein. Hingegen priorisieren Address-Scans der USA hauptsächlich das zuvor angesprochene *Session Initiation Protocol*. Address-Scans, welche von Großbritannien durchgeführt werden, verwenden fast keine der im Diagramm aufgeführten Protokolle. Weshalb das Scan-Verhalten der USA so einseitig ausfällt, lässt sich an dieser Stelle nicht weiter begründen. Gleiches gilt für das zu Großbritannien beschriebene Scan-Verhalten. Wie schon zuvor, wäre auch hier eine Gegenprobe mittels Daten eines anderen Analysezeitraums sinnvoll.

Die Ergebnisse der beiden Diagramme 5.12 und 5.13 werden im Rahmen des folgenden Abschnitts genauer interpretiert. Das Scan-Verhalten vieler der dort aufgeführten RIR-Namen entspricht den in 5.7 und 5.8 erkannten Phänomenen. So ist auf *ssd networks limited* in 5.12 die zweithöchste Anzahl an Address-Scans zurückzuführen. Für Port-Scans ist hingegen nur eine geringe Anzahl an Scans, aber eine enorme Anzahl an Scan-Anfragen zu verzeichnen. Dieses Scan-Verhalten lässt sich in den Ausreißern der beiden genannten Abbildungen wiederfinden. So ist für *ssd networks limited* in 5.7 ausschließlich ein Hochpunkt der Address-Scans zu verzeichnen. Bei Betrachtung der tatsächlichen Scan-Anfragen in 5.8 ist für denselben Zeitpunkt ein Ausreißer beider Scan-Typen zu erkennen. Durch den Zusammenhang dieser unterschiedlichen Ab-

bildungen lässt sich die zeitliche Komponente aus 5.8 auf die Diagramme 5.12 und 5.13 übertragen. So kann vermuten werden, dass ein großer Anteil der zu *ssd networks limited* in 5.12 und 5.13 visualisierten Scans in genau dem Zeitraum der Ausreißer durchgeführt wurden. Weiter zu diesem RIR-Namen in 5.12 festzustellen ist, dass dieser die größte Anzahl unterschiedlicher Ziel-Ports zu verantworten hat. Diese Informationen kombiniert mit der aufgestellten Vermutung lässt darauf schließen, dass von *ssd networks limited* im genannten Zeitraum nicht nur eine enorme Anzahl aggressiver Scans ausgegangen ist, sondern dass diese Scans auch auf eine große Anzahl unterschiedlicher Ziel-Ports gerichtet waren. Das beschriebene Scan-Verhalten von *ssd networks limited* könnte durch die Sicherheitsmechanismen der Cloud-Anbieter erklärt werden. Firewalls dieser Cloud-Anbieter überwachen große IP-Adressbereiche. Groß angelegte Address-Scans, welche zudem eine Vielzahl unterschiedlicher Ziel-Ports abfragen, sind von Firewalls leichter zu erkennen, wenn diese nur von einer oder wenigen Quell-IPs ausgehen[41]. Die hohe Anzahl einzeln identifizierter Address-Scans könnte somit der Versuch sein, die Scan-Anfragen auf verschiedene Quell-IPs zu verteilen.

Eine ähnliche Verknüpfung der beiden Diagramme lässt sich mit dem Scan-Verhalten von *DigitalOcean, LLC* bzw. der zugehörigen IP *188.166.250.148* herstellen. Wie in Unterabschnitt 5.2.4 zu den Ergebnissen des Diagramms beschrieben, ist diese IP auch dort für einen großen Anteil der Balken von *DigitalOcean, LLC* verantwortlich. Bei Betrachtung der Address-Scans ist für diesen RIR-Namen die weitaus höchste Anzahl an Scan-Anfragen bei einer vergleichbar geringen Anzahl von Scans zu verzeichnen. Entsprechend ist bei Betrachtung der beiden Koordinatensysteme 5.7 und 5.8 festzustellen, dass für *DigitalOcean, LLC* ausschließlich Ausreißer bei dem Verlauf der tatsächlichen Scan-Anfragen zu erkennen sind. Bei dem Verlauf der Anzahl identifizierter Scans sind solche Ausreißer nicht festzustellen.

Wie weiter in Unterabschnitt 5.2.4 beobachtet, ist für *NSEC - Sistemas Informaticos, S.A.* eine sehr hohe Scandauer bei vergleichsweise wenigen Scans und Anfragen zu erkennen. Ein solches Scan-Verhalten kann wieder durch vorherrschende Sicherheitsmechanismen begründet werden. So ist es schwieriger, langsame Scans mit einer geringen Paket-Rate von legitimen Traffic zu unterscheiden[41].

6.2.2 Bewertung DoS/DDoS-Angriffe

Zu Beginn dieses Unterkapitels werden die in 5.15 beschriebenen Ergebnisse genauer interpretiert. Das Diagramm veranschaulicht, welche Protokolle bei den aufgezeichneten Angriffen am häufigsten verwendet wurden. Durch einen Vergleich mit den gescannten Ports, welche in 5.15 visualisiert sind, kann eine Begründung für viele der angegriffenen Protokolle gefunden werden. So verzeichnen vor allem die Protokolle *DNS* und *Session Initiation Protocol* in beiden Diagrammen hohe Werte. Dies entspricht genau der in Abschnitt 3.6

dargestellten Feststellung, wonach vor der Durchführung von Angriffen häufig genaue Scans des Ziels stattfinden. Wie dort weiter beschrieben, werden diese Scans durchgeführt, um Informationen über offene/verwundbare Ports herauszufinden. Die hohe Anzahl an Angriffen, welche auf das *Session Initiation Protocol* abgezielt haben, lässt sich zudem mit den im vorigen Kapitel beschriebenen Sicherheitsrisiken erklären.

Nachstehend erfolgt eine Diskussion über den geografischen Zusammenhang zwischen den angegriffenen Ländern 5.16 und dem Ursprung dieser Angriffe 5.17. Hierzu eignet sich ein Vergleich mit den Ergebnissen des Papers von Pauley *et al.*[1]. Das aufgeführte Paper eignet sich für diesen Vergleich, da dort ein ähnliches Cloud-Teleskop zum Aufzeichnen von Internet-Traffic verwendet wird. In der Arbeit wird eine ähnliche geografische Betrachtung zum Thema Scanning durchgeführt. Hierzu werden, ähnlich wie in 5.17, die unterschiedlichen Regionen des Cloud-Teleskops miteinander verglichen. So wird verglichen, in welchem Ausmaß sich die von Scannern genutzten src-IPs in den jeweiligen Regionen überschneiden. Weiter wird analysiert, ob benachbarte Regionen eine höhere Überlappung gleicher src-IPs aufweisen als weiter voneinander entfernte Regionen. Das Ergebnis dieses Vergleiches ist, dass die geografische Lage des gescannten Hosts keinen nennenswerten Zusammenhang mit dem Ursprung des Scanners aufweist. Mithilfe von 5.17 wurde in dieser Arbeit ein ähnlicher Vergleich in Bezug auf DoS/DDoS-Angriffe durchgeführt. Genauer wurde ein Zusammenhang zwischen der geografischen Lage der gefälschten src-IP und dem angegriffenen Land untersucht. Wie im Ergebnisteil dargestellt, ist auch für DoS/DDoS-Angriffe kein solcher geografischer Zusammenhang festzustellen. Die im Paper herausgefundenen Ergebnisse zum Thema Scanning lassen sich somit auf den Bereich der DoS/DDoS-Attacken übertragen.

Durch eine Interpretation der zu 5.18 gefundenen Ergebnisse lassen sich interessante Vermutungen aufstellen. In dem Diagramm sind die angegriffenen RIR-Namen aufgeführt, von denen die meisten Backscatter-Antworten empfangen wurden. Die Tatsache, dass unter diesen RIR-Namen auch große Technologieunternehmen wie Microsoft oder Google zu finden sind, ist auf die enorme Internetpräsenz dieser Firmen zurückzuführen. Eine Vielzahl der aufgeführten DNS-Reverse-Einträge kennzeichnet gemietete Cloud-Maschinen von mittelgroßen bis großen Cloud-Anbietern. Die Erkenntnis, dass viele dieser Cloud-Anbieter aus den USA stammen, lässt sich bei der Betrachtung der angegriffenen Länder in 5.16 wiederfinden. Durch die Ergebnisse der nmap-Analyse können nun weitere Vermutungen aufgestellt werden.

Wie zu dem Diagramm beschrieben, konnte zu einigen Servern mittels nmap überhaupt keine Verbindung aufgebaut werden. Dies kann entweder durch sehr restriktive Firewall-Einstellungen begründet werden oder dadurch, dass der Server abgestellt ist bzw. nicht existiert. Sollte ersteres gelten, so lassen sich die im Teleskop registrierten Backscatter-Antworten nicht erklären, da die entspre-

chende Maschine dann auch nicht auf die vom Angreifer gesendeten Anfragen geantwortet hätten. Eine Schlussfolgerung könnte sein, dass möglicherweise der Angriff selbst zu einer Anpassung der Schutzmechanismen geführt hat.

Zu den anderen Cloud-Maschinen können die gefundenen Ports Hinweise auf die Verwendung des angegriffenen Servers geben. So kann anhand der beschriebenen Ergebnisse vermutet werden, dass unter *vmi1278843.contaboserver.net.* ein DNS-Server läuft. Für *v281302.hosted-by-vdsina.com.* sprechen die gefundenen Ports für das Hosting einer Internetseite. Die Ports 135-139 deuten hierbei zudem auf ein Microsoft Betriebssystem hin[54]. Weiter ist festzustellen, dass bei vielen Cloud-Maschinen die offenen Ports mit denen übereinstimmen, welche bei dem jeweiligen Angriff verwendet wurden. So konnte zu *216-10-250-164.webhostbox.net.* der offene Port für das *Session Initiation Protocol* gefunden werden. Dieses Protokoll stimmt mit allen von dieser IP registrierten Backscatter-Antworten überein, da diese ausschließlich an Port 5060 gerichtet sind.

Abschließend folgt eine kurze Auseinandersetzung mit dem Ergebnis zur Analyse der Ausfallzeiten angegriffener Systeme. Das hierfür verwendete ICMP-Szenario konnte nur für einen einzigen Angriff nachgewiesen werden. Diese sehr geringe Anzahl lässt sich durch mehrere Gründe erklären. Zunächst einmal führt nicht jeder Angriff automatisch zur erfolgreichen Unterbrechung des Zielservers. Dies kann an entsprechenden Abwehrmechanismen oder einem zu klein dimensionierten Angriff liegen. Weiter muss der Angriff, wie in Abschnitt 3.5 beschrieben, neben UDP auch mittels TCP erfolgen, damit der angegriffener Server entsprechende ICMP-Pakete versendet. Alle Angriffe, welche zwar erfolgreich durchgeführt wurden, aber ausschließlich auf UDP basieren, werden somit nicht betrachtet. Weiter liegt es an den Konfigurationen des jeweiligen Servers, ob dieser bei einer Überlastung die notwendigen ICMP-Antworten versendet. Wie schon in Abschnitt 3.5 angedeutet, kann auch die begrenzte Mietdauer der jeweiligen Maschinen des Cloud-Teleskops dazu führen, dass ein Angriffsverlauf nicht vollständig erfasst wird.

Somit kann gesagt werden, dass das Einbeziehen beider Transportprotokolle vermutlich zu aussagekräftigeren Ergebnissen führen würde. Eine solche erneute Überprüfung des beschriebenen ICMP-Szenarios könnte in einer folgenden Arbeit umgesetzt werden.

6.3 Bewertung der im CAIDA-Teleskop wiedergefundenen Phänomene

Nachstehend erfolgt nun eine Diskussion der in Abschnitt 5.4 vorgestellten Ergebnisse über den Vergleich der beiden Teleskope. Diese Diskussion beginnt mit einer genaueren Interpretation der zu 5.19a und 5.19b aufgestellten Beobachtungen. Wie der Vergleich beider Abbildungen zeigt, ist der prozentuale

Anteil an Address-Scans, welche im CAIDA-Teleskop gefunden wurden, deutlich höher als der Anteil der Port-Scans. Anhand der weiter beschriebenen Ergebnisse ist zu erkennen, dass diese Beobachtung nicht durch einen geografischen Zusammenhang zu erklären ist. So kann für beide Scan-Typen nachgewiesen werden, dass der Ursprung der Scans in keiner nennenswerten Weise mit der geografischen Lage des Teleskops korreliert. Anhand dieser beiden Erkenntnisse lässt sich schlussfolgern, dass die Ziele der Port-Scans im Gegensatz zu denen der Address-Scans nicht zufällig ausgesucht werden. Weiter bedeutet dies, dass viele Port-Scans den IP-Adressbereich des CAIDA-Teleskops meiden. Somit kann gesagt werden, dass die Ziele der Port-Scans deutlich sorgfältiger ausgesucht werden als die der Address-Scans. Eine Begründung hierfür könnte sein, dass der Scanvorgang eines Port-Scans für jede Ziel-IP deutlich mehr Zeit in Anspruch nimmt. Die Erklärung hierfür liegt in der Definition eines Port-Scans. So führt dieser, im Gegensatz zu einem Address-Scan, mehrere Scan-Anfragen für jedes Ziel durch.

Ein Vergleich der RIR-Namen aus 5.20 mit den zuvor in Abschnitt 6.1 aufgestellten Vermutungen liefert weitere interessante Erkenntnisse. So sind in den beiden Tabellen aus 5.20 jeweils die RIR-Namen aufgeführt, welche ausschließlich im Cloud-Teleskop gefunden wurden, und die RIR-Namen welche auch im CAIDA-Teleskop nachgewiesen wurden. Bei der Betrachtung ersterer ist schnell festzustellen, dass viele dieser RIR-Namen schon in der Diskussion des vorherigen Kapitels thematisiert wurden. Vor allem ein Vergleich mit den RIR-Namen aus 5.8, welche dort für die Ausreißer verantwortlich sind, lässt eine interessante Vermutung aufstellen. So lässt sich an diesem Vergleich erkennen, dass viele der nur in den Daten des Cloud-Teleskops gefundenen Scans genau die Scans sind, welche zuvor durch ein sehr aggressives Scan-Verhalten mit einer enormen Menge an Anfragen aufgefallen sind. Dies lässt die begründete Vermutung aufstellen, dass gerade diese sehr aggressiven Scans die IP-Adressen des CAIDA-Teleskops meiden.

Weiter lässt sich erkennen, dass die Scans der aus dem Iran stammenden RIR-Namen *Mobile Communication Company of Iran PLC*, *Information Technology Company (ITC)* und *Iran Telecommunication Company PJS* nicht in den CAIDA-Daten gefunden wurden. Zu diesen RIR-Namen konnte in Abschnitt 6.1 die Vermutung aufgestellt werden, dass der Betreiber zur Durchführung von Address-Scans eine Cloud-Maschine von Digital-Ocean verwendet hat. Die Erkenntnis, dass diese Scans nicht im CAIDA-Teleskop gefunden wurden, kombiniert mit dem Fakt, dass genau von dem Cloud-Anbieter eine VM gemietet wurde, welcher auch für das verwendete Cloud-Teleskop ausgesucht wurde, lässt vermuten, dass es die Scans des Betreibers ausschließlich auf die Digital-Ocean-Infrastruktur gerichtet waren. Ansonsten wäre es ein starker Zufall, dass kein Zusammenhang vorliegt zwischen der enormen Menge an Scan-Anfragen, welche an IP-Adressen des Digital-Ocean IP-Adressbereichs gerichtet sind und der für diese Scans gemieteten Cloud-Maschinen, welche

vom selben Cloud-Anbieter stammen. Um diese Vermutung genauer zu validieren, wären zu dem Analysezeitraum Daten eines weiteren Cloud-Teleskops notwendig, welches auf IP-Adressen eines anderen Cloud-Anbieters basieren. Sollten sich diese Vermutungen allerdings bewahrheiten, so konnte gezeigt werden, dass entsprechende Betreiber aus dem Iran gezielt Cloud-Maschinen des jeweiligen Cloud-Anbieters mieten, um mittels dieser Maschinen aggressive Address-Scans der jeweiligen Cloud-Infrastruktur durchzuführen.

Der weitere Teil dieser Diskussion beschäftigt sich nun mit dem Vergleich der gefundenen DoS/DDoS-Angriffe beider Teleskope. Anhand der Ergebnisse aus 5.21 ist zu erkennen, dass weniger als die Hälfte aller Angriffe im CAIDA-Teleskop gefunden wurden. Diese Beobachtung widerspricht zunächst den Erkenntnissen aus 5.17, wo gezeigt wurde, dass der Ursprung der für einen Angriff verwendeten src-IPs keinen geografischen Zusammenhang mit dem angegriffenen Ziel aufweist. Durch den Vergleich der Anzahl an Backscatter-Verbindungen in 5.21 kann allerdings festgestellt werden, dass die geringe Anzahl der im CAIDA-Teleskop gefundenen Angriffe einen erheblichen Anteil der Backscatter-Antworten des Cloud-Teleskops zu verantworten hat. Dies lässt darauf schließen, dass vor allem große Angriffe in den Daten des CAIDA-Teleskops wiedergefunden wurden. Da ein geografischer Zusammenhang ausgeschlossen wurde, kann für die Anfragen der kleineren Angriffe vermutet werden, dass diese auf Fehlkonfigurationen zurückzuführen sind und so keine Angriffe darstellen. Um diese vermeintlichen Angriffe herauszufiltern, wäre eine weiterführende Analyse unter Einbeziehung eines entsprechenden Grenzwertes an Backscatter-Antworten notwendig. Ein Vergleich, ab welcher Anzahl registrierter Backscatter-Antworten im Cloud-Teleskop ein Angriff auch in den Daten des CAIDA-Teleskops nachzuweisen ist, könnte für die Bestimmung eines solchen Grenzwertes herangezogen werden.

In den Ergebnissen zu 5.22 sind viele der zuvor in Abschnitt 6.1 diskutierten Angriffe wiederzufinden. So konnten auch die zuvor genauer untersuchten Cloud-Maschinen *vmi1278843.contaboserver.net.*, *v281302.hosted-by-vdsina.com.* und *216-10-250-164.webhostbox.net.* in den Daten des CAIDA-Teleskops gefunden werden. Somit dient das erneute Auffinden dieser Angriffe zur Untermauerung der Authentizität dieser Identifizierungen. Die enormen Unterschiede in der Anzahl der zu den Angriffen erfassten Antworten lassen sich durch die Dimensionierung der beiden Teleskope begründen. So deckt das CAIDA-Teleskop ein Vielfaches mehr an IP-Adressen ab, als die Anzahl an IP-Adressen, welche vom Cloud-Teleskop im Analysezeitraum verwendet wurden. Aufgrund dieses Größenunterschieds lassen sich auch die unterschiedlichen Positionierungen der RIR-Namen in beiden Diagrammen begründen. Um vergleichbare Ergebnisse einer solchen Gegenüberstellung zu erzielen, wäre eine solche Dimensionierung des Cloud-Teleskops notwendig, sodass die Anzahl der verwendeten IP-Adressen der des Darknet-Teleskops entspricht.

7 Zusammenfassung

Der folgende Abschnitt fasst die zuvor gewonnenen Erkenntnisse zusammen. Zunächst wurden einige Verfahren vorgestellt, um die zur Erkennung von Scans notwendigen Parameter eines minimalen Anfrage-Grenzwerts sowie eines zeitlichen Intervalls zu ermitteln. Zur Bestimmung des Anfrage-Grenzwerts wurden drei Analysen durchgeführt, welche einen Verlauf in Abhängigkeit von unterschiedlichen Grenzwerten und Intervallwerten visualisieren. Hierbei wurden die Verläufe der Anzahl gefundener Scans, der Anzahl neu gefundener Scanner sowie der durchschnittlichen Scandauer betrachtet. Durch eine Sensitivitätsanalyse konnte ein entsprechender Anfragen-Grenzwert festgelegt werden. Daran anschließend wurde ein Verfahren vorgestellt, welches die normalisierten Funktionswerte der zuvor beschriebenen Verläufe darstellt. Der durch die Normalisierung entstandene Zusammenhang der Funktionsverläufe wurde genutzt, um ein entsprechendes Intervall zu ermitteln. Anschließend wurden eine Reihe von Analysen vorgestellt, um spezielle Phänomene in den Daten des Cloud-Teleskops aufzudecken. So konnte nachgewiesen werden, dass einzelne RIR-Namen extrem aggressive Scans zu verantworten haben. Bei der geografischen Betrachtung des Scan-Ursprungs konnte am Beispiel von China durch einen Vergleich mit einer anderen Studie die Vermutung aufgestellt werden, dass sich die Scanstrategien der einzelnen Länder zwischen TCP und UDP unterscheiden. Bei der genaueren Betrachtung einzelner Scans konnte gezeigt werden, dass einige Scan-Verhalten darauf abzielen, Sicherheitsmechanismen zu umgehen. Genauer wurde für einige Scans festgestellt, dass Anfragen auf viele einzelne Address-Scans aufgeteilt werden oder dass sehr langsame Scan-Raten verwendet werden. Bei der Analyse verübter DoS/DDoS-Angriffe konnte gezeigt werden, dass für diese Attacken hauptsächlich die Protokolle DNS und SIP verwendet werden. Weiter konnte in einer umfassenden Analyse gezeigt werden, dass der Ursprung der für einen Angriff verwendeten src-IPs keinen geografischen Zusammenhang zum angegriffenen Ziel aufweist. Bei einer genaueren Betrachtung der attackierten Ziele ließen sich anhand der durch eine nmap-Analyse identifizierten Ports Rückschlüsse auf den Verwendungszweck der Server ziehen. Durch einen Vergleich mit den Daten des CAIDA-Teleskops konnte gezeigt werden, dass vor allem Port-Scans die IP-Adressen dieses Teleskops meiden. Weiter wurde nachgewiesen, dass gerade die aggressiven Scans nicht in den CAIDA-Daten wiederzufinden sind. Abschließend konnte die begründete Vermutung aufgestellt werden, dass entsprechende Betreiber aus dem Iran gezielt Cloud-Maschinen eines Cloud-Anbieters mieten, um mittels dieser Maschinen Address-Scans der jeweiligen Cloud-Infrastruktur durchzuführen.

Abbildungsverzeichnis

5.1 Prozentuale Verteilung der Transportprotokolle. 21
5.2 Prozentuale Verteilung gefundener Anwendungsprotokolle. . . . 21
5.3 Anzahl gefundener Scans in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls. 23
5.4 Anzahl neu gefundener Scanner in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls. 24
5.5 Durchschnittliche Scandauer in Abhängigkeit des Anfragen-Grenzwerts sowie des zeitlichen Intervalls. 25
5.6 Normalisierte Funktionsverläufe der Anzahl gefundener Scans, der Anzahl neu gefundener Scanner und der durchschnittlichen Scandauer bei einem Anfragen-Grenzwert von 6. 26
5.7 Zeitlicher Verlauf identifizierter Port- und Address-Scans (60-Minuten-Aggregation). 28
5.8 Zeitlicher Verlauf registrierter Anfragen der Port- und Address-Scans (60-Minuten-Aggregation). 29
5.9 Verteilung der identifizierten Scans auf die Top 20 Herkunftsländer sowie die Einteilung anhand der Scan-Typen. 30
5.10 Prozentuale Verteilung der Top 20 am meisten gescannten Protokolle. 30
5.11 Verteilung der Scans auf die Top 10 am meisten gescannten Protokolle, differenziert nach den Scan-Herkunfsländern. 31
5.12 Address-Scans: Vergleich der Top zehn RIR-Namen anhand der Anzahl an Scans, der Anzahl an Scan-Anfragen, der durchschnittlichen Scandauer und der Anzhal unterschiedlich gescannter Ziel-Ports. 32
5.13 Port-Scans: Vergleich der Top zehn RIR-Namen anhand der Anzahl an Scans, der Anzahl an Scan-Anfragen, der durchschnittlichen Scandauer und der Anzhal unterschiedlich gescannter Ziel-IPs. 33
5.14 Übersicht der zu Forschungszwecken betriebenen Scans. 34
5.15 Verteilung der Angriffe anhand der verwendeten Portokolle. . . . 35
5.16 Verteilung der Angriffe auf die Top 20 am stärksten attackierten Länder. 35
5.17 Anzahl gefundener Angriffe in Abhängigkeit des attackierten Landes sowie des geografischen Ursprungs der für den Angriff verwendeten src-IP. 36

5.18 Verteilung der registrierten Backscatter-Antworten auf die Top 20 am stärksten attackierten RIR-Namen. 38
5.19 Vergleich erkannter Scanner beider Teleskope sowie deren geografische Einordnung. 40
5.20 Übersicht der Top 10 RIR-Namen beider Teleskope, welche die meisten Scans zu verantworten haben. 41
5.21 Vergleich beider Teleskope anhand der Anzahl gefundener Angriffe sowie der Anzahl registrierter Backscatter-Antworten. . . . 42
5.22 Verteilung der im CAIDA-Teleskop registrierten Backscatter-Antworten auf die Top 20 am stärksten attackierten RIR-Namen. 43

Literatur

1. PAULEY, Eric; BARFORD, Paul; MCDANIEL, Patrick. {DScope}: A {Cloud-Native} Internet Telescope. In: *32nd USENIX Security Symposium (USENIX Security 23)*. 2023, S. 5989–6006. ISBN 978-1-939133-37-3.
2. SPACELIFT. *55 Cloud Computing Statistics for 2025* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://spacelift.io/blog/cloud-computing-statistics`.
3. ADIVI CORPORATION. *40+ Cloud Security Statistics You Need to Know in 2024* [online]. 2024. [besucht am 2025-03-30]. Abger. unter: `https://adivi.com/blog/cloud-security-statistics/`.
4. MOORE, David; SHANNON, Colleen; BROWN, Douglas J; VOELKER, Geoffrey M; SAVAGE, Stefan. Inferring internet denial-of-service activity. *ACM Transactions on Computer Systems (TOCS)*. 2006, Jg. 24, Nr. 2, S. 115–139. Abger. unter DOI: `10.1145/1132026.1132027`.
5. MOORE, David; SHANNON, Colleen; VOELKER, Geoffrey M.; SAVAGE, Stefan. *Network Telescopes: Technical Report*. 2004. Techn. Ber., CS2004-0795. Department of Computer Science & Engineering, UC San Diego.
6. BLENN, Norbert; GHIËTTE, Vincent; DOERR, Christian. Quantifying the Spectrum of Denial-of-Service Attacks through Internet Backscatter. In: *Proceedings of the 12th International Conference on Availability, Reliability and Security*. ACM, 2017, S. 1–10. Abger. unter DOI: `10.1145/3098954.3098985`.
7. ROSSOW, Christian. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. In: *Proceedings 2014 Network and Distributed System Security Symposium*. Internet Society, 2014, S. 1–15. Abger. unter DOI: `10.14722/ndss.2014.23233`.
8. DURUMERIC, Zakir; BAILEY, Michael; HALDERMAN, J. Alex. An {Internet-Wide} View of {Internet-Wide} Scanning. In: *23rd USENIX Security Symposium (USENIX Security 14)*. 2014, S. 65–78. ISBN 978-1-931971-15-7.
9. BHUYAN, Monowar H; BHATTACHARYYA, Dhruba Kr; KALITA, Jugal K. Surveying Port Scans and Their Detection Methodologies. *The Computer Journal*. 2011, Jg. 54, Nr. 10, S. 1565–1581. Abger. unter DOI: `10.1093/comjnl/bxr035`.

10. HERCOG, Drago. *Communication Protocols: Principles, Methods and Specifications*. Cham: Springer International Publishing, 2020. Abger. unter DOI: DOI:10.1007/978-3-030-50405-2.
11. BORDEL, Stefan. *Was ist das OSI-Modell?* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://www.myrasecurity.com/en/knowledge-hub/osi-modell/.
12. MOCKAPETRIS, Paul. *DOMAIN NAMES - CONCEPTS AND FACILITIES* [online]. 1987. [besucht am 2025-03-30]. Abger. unter: https://www.rfc-editor.org/rfc/rfc1034.
13. SRĖBALIŪTĖ, Agnė. *Public vs. private DNS servers* [online]. 2024. [besucht am 2025-03-30]. Abger. unter: https://nordlayer.com/blog/public-vs-private-dns-servers/.
14. MOCKAPETRIS, Paul. *DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION* [online]. 1987. [besucht am 2025-03-30]. Abger. unter: https://www.rfc-editor.org/rfc/rfc1035.
15. ARENDS, Rob; AUSTEIN, Rodney; LARSON, Matt; MASSEY, Dave; ROSE, Steve. *Resource Records for the DNS Security Extensions* [online]. 2005. [besucht am 2025-03-30]. Abger. unter: https://www.rfc-editor.org/rfc/rfc4034.
16. CHESHIRE, Stuart; KROCHMAL, Marc. *Multicast DNS* [online]. 2013. [besucht am 2025-03-30]. Abger. unter: https://datatracker.ietf.org/doc/html/rfc6762.
17. BUNDESAMT FÜR SICHERHEIT IN DER INFORMATIONSTECHNIK (BSI). *Offene mDNS-Dienste* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Cyber-Sicherheitslage/Reaktion/CERT-Bund/CERT-Bund-Reports/HowTo/Offene-mDNS-Dienste/Offene-mDNS-Dienste_node.html?utm_source=chatgpt.com.
18. CYBERSECOURITY INFRASTRUCTURE SECURITY AGENCY. *UDP-Based Amplification Attacks* [online]. 2019. [besucht am 2025-03-30]. Abger. unter: https://www.cisa.gov/news-events/alerts/2014/01/17/udp-based-amplification-attacks.
19. MILLS, David. *Network Time Protocol (Version 3) Specification, Implementation and Analysis* [online]. 1992. [besucht am 2025-03-30]. Abger. unter: https://datatracker.ietf.org/doc/html/rfc1305.
20. THE DEBIAN PROJECT. *NTPD(8)* [online]. 2018. [besucht am 2025-03-30]. Abger. unter: https://manpages.debian.org/buster/ntp/ntpd.8.en.html.
21. CLOUDFLARE. *NTP-Amplification-DDoS-Angriff* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://www.cloudflare.com/learning/ddos/ntp-amplification-ddos-attack/.

22. IRWIN, Barry. *A framework for the application of Network Telescope Sensors in a global IP Network*. Rhodes University, 2011. Abger. unter DOI: 10.23721/107/1421850. Diss.

23. CAIDA. *Sustainable Tools for Analysis and Research on Darknet Unsolicited Traffic* [online]. 2021. [besucht am 2025-03-30]. Abger. unter: https://www.caida.org/projects/stardust/.

24. GU, Qijun; LIU, Peng. Denial of Service Attacks. *Handbook of Computer Networks: Distributed Networks, Network Planning, Control, Management, and New Trends and Applications*. 2007, Jg. 3, S. 454–468. Abger. unter DOI: 10.1002/9781118256107.ch29.

25. TORABI, Sadegh; BOU-HARB, Elias; ASSI, Chadi; GALLUSCIO, Mario; BOUKHTOUTA, Amine; DEBBABI, Mourad. Inferring, Characterizing, and Investigating Internet-Scale Malicious IoT Device Activities: A Network Telescope Perspective. In: *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 2018, S. 562–573. Abger. unter DOI: 10.1109/DSN.2018.00064.

26. ZEEK. *About Zeek* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://docs.zeek.org/en/master/about.html#.

27. ZEEK. *Packet Analysis* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://docs.zeek.org/en/master/frameworks/packet-analysis.html.

28. ZEEK. *Zeek Script Index* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://docs.zeek.org/en/master/script-reference/scripts.html.

29. HALL, Seth; SHARMA, Aashish. *scan_udp.bro* [online]. 2013. [besucht am 2025-03-30]. Abger. unter: https://github.com/sethhall/zeek-junk-drawer/blob/master/scan_udp.bro.

30. UNIVERSITÄT MÜNSTER. *Prof. Dr. Ralph Holz, Institut für Informatik* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://www.uni-muenster.de/Informatik/organisation/show_perspage.shtml?id=1746.

31. DIGITAL OCEAN. *The simplest cloud that scales with you* [online]. [besucht am 2025-03-30]. Abger. unter: https://www.digitalocean.com.

32. CAIDA. *The UCSD Network Telescope* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: https://www.caida.org/projects/network_telescope/.

33. CAIDA. *Traffic Traces (pcap)* [online]. 2021. [besucht am 2025-03-30]. Abger. unter: https://www.caida.org/projects/stardust/docs/data/pcap/.

34. ALCOCK, Shane. *Flowtuples IV: Reality Strikes Back* [online]. 2021. [besucht am 2025-03-30]. Abger. unter: `https://www.caida.org/catalog/media/2021_flowtuples_iv_dust/flowtuples_iv_dust.pdf`.

35. CAIDA. *Libtrace - STARDUST* [online]. 2021. [besucht am 2025-03-30]. Abger. unter: `https://www.caida.org/projects/stardust/docs/tutorials/libtracetutorial/`.

36. LIBTRACETEAM. *tracepktdump* [online]. 2014. [besucht am 2025-03-30]. Abger. unter: `https://github.com/LibtraceTeam/libtrace/wiki/tracepktdump`.

37. LIBTRACETEAM. *Filter Expressions* [online]. 2014. [besucht am 2025-03-30]. Abger. unter: `https://github.com/LibtraceTeam/libtrace/wiki/Filter-Expressions`.

38. KEMPEN, Nils. *Cloud-based Internet Telescope* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/thisni1s/telescope`.

39. JUTZI, Tobias. *Bachelor-Thesis* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/tobiasjutzi/Bachelor-Thesis/blob/main/Cloud-Telescope/notebooks/11_scan_validation.ipynb`.

40. AZOFF, Justin; SIWEK, Jon; dopheide esnet dopheide; AVILA, Kay; CULLEN, Peter. *Bro simple scan* [online]. 2022. [besucht am 2025-03-30]. Abger. unter: `https://packages.zeek.org/packages/view/ce3d0fe0-9348-11eb-81e7-0a598146b5c6`.

41. RICHTER, Philipp; BERGER, Arthur. Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope. In: *Proceedings of the Internet Measurement Conference* [online]. ACM, 2019, S. 144–157 [besucht am 2025-03-15]. Abger. unter DOI: `10.1145/3355369.3355595`.

42. ZEEK. *base/protocols/conn/main.zeek* [online]. 2024. [besucht am 2025-03-30]. Abger. unter: `https://docs.zeek.org/en/lts/scripts/base/protocols/conn/main.zeek.html`.

43. JUTZI, Tobias. *Bachelor-Thesis* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/tobiasjutzi/Bachelor-Thesis/blob/main/Cloud-Telescope/notebooks/12_backscatter_validattion.ipynb`.

44. CRAWFORD, Dave; AZOFF, Justin. *Connection History: connection direction was flipped by Bro's heuristic* [online]. 2017. [besucht am 2025-03-30]. Abger. unter: `https://community.zeek.org/t/connection-history-connection-direction-was-flipped-by-bro-s-heuristic/4798`.

45. JUTZI, Tobias. *ipInfo-script.sh* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/tobiasjutzi/Bachelor-Thesis/blob/main/CAIDA-Telescope/scripts/ipInfo-script.sh`.

46. IPINFO. *The Trusted Source For IP Address Data* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://ipinfo.io`.

47. SHAHZAD, Uman; CURIONI, Maxime; FAROOQ, Umar; SHAHZAD, Shamir; USAMA, Abu; MOUCHET, Maxime; ST, Polina; MURAD, Muhammad; DEVREL, Abdullah. *mmdbctl* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/ipinfo/mmdbctl?ref=ipinfo.io`.

48. JUTZI, Tobias. *script_merge_geo_databasey* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/tobiasjutzi/Bachelor-Thesis/blob/main/Cloud-Telescope/src/d02_intermediate/script_merge_geo_database.py`.

49. RÉSEAUX IP EUROPÉENS NETWORK COORDINATION CENTRE. *Regional Internet Registry* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://www.ripe.net/about-us/what-we-do/regional-internet-registry/`.

50. JUTZI, Tobias. *Bachelor-Thesis* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://github.com/tobiasjutzi/Bachelor-Thesis`.

51. NEWMAN, Sean. *What is CoAP (Constrained Application Protocol)?* [online]. 2024. [besucht am 2025-03-30]. Abger. unter: `https://www.corero.com/what-is-constrained-application-protocol/`.

52. WHITE, Steven; CAI, Saisang; PETERSEN, Theano; SHARKEY, Kent; COULTER, David; JACOBS, Mike; SATRAN, Michael. *About Web Services on Devices* [online]. 2021. [besucht am 2025-03-30]. Abger. unter: `https://learn.microsoft.com/en-us/windows/win32/wsdapi/about-web-services-for-devices`.

53. LYON, Gordon. *Nmap Reference Guide* [online]. 2009. [besucht am 2025-03-30]. Abger. unter: `https://nmap.org/book/man.html`.

54. LIANG, Han; XU, Simonx; LI, Anna; PHOEBE; APPELGATE, Teresa; MANAGAOK0314; HUNGTSETSE; MISASA0818; SIMPSON, Robert; SMETS, Martine; QIU, Marry. *Service overview and network port requirements for Windows* [online]. 2025. [besucht am 2025-03-30]. Abger. unter: `https://learn.microsoft.com/en-us/troubleshoot/windows-server/networking/service-overview-and-network-port-requirements`.

Eidesstattliche Erklärung

Hiermit versichere ich, dass die vorliegende Arbeit über *„Vergleich von klassischen und Cloud-basierten Netzwerk-Teleskopen anhand von UDP- und ICMP-Paketen“* selbstständig verfasst worden ist, dass keine anderen Quellen und Hilfsmittel als die angegebenen benutzt worden sind und dass die Stellen der Arbeit, die anderen Werken – auch elektronischen Medien – dem Wortlaut oder Sinn nach entnommen wurden, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht worden sind.

Vorname Nachname, Münster, 31. März 2025

Ich erkläre mich mit einem Abgleich der Arbeit mit anderen Texten zwecks Auffindung von Übereinstimmungen sowie mit einer zu diesem Zweck vorzunehmenden Speicherung der Arbeit in eine Datenbank einverstanden.

Vorname Nachname, Münster, 31. März 2025