HAW HAMBURG

BACHELORARBEIT

Vergleich von DDoS-Angriffstypen und deren Erkennung durch KI-gestützte Systeme

vorgelegt am 18. September 2025

Mursal Kul

Erstprüfer: Prof. Dr. Nils Martini
Zweitprüfer: Prof. Dr. Edmund Weitz

HOCHSCHULE FÜR ANGEWANDTE WISSENSCHAFTEN HAMBURG
Department Medientechnik
Finkenau 35
20081 Hamburg

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Erkennung von Distributed Denial of Service (DDoS)-Angriffen durch den Einsatz von Methoden der Künstlichen Intelligenz (KI). Angesichts der zunehmenden Häufigkeit und Komplexität solcher Angriffe stoßen klassische Abwehrmechanismen wie Firewalls und Intrusion Detection Systeme an ihre Grenzen. Ziel der Arbeit ist es, verschiedene Machine-Learning- (ML) und Deep-Leraning (DL) -Ansätze zu analysieren und ihre Wirksamkeit bei der Erkennung unterschiedlicher Angriffstypen zu vergleichen.

Im theoretischen Teil werden die Grundlagen von DDoS-Angriffen, ihre Typen und Auswirkungen erläutert sowie die Funktionsweise von ML- und DL-Ansätzen dargestellt. Anschließend erfolgt eine vergleichende Analyse der Modellleistungen anhand von Studienergebnissen. Dabei zeigt sich, dass ML-Modelle wie Random Forest oder SVM durch Effizienz und einfache Implementierung überzeugen, während DL-Modelle wie CNNs oder LSTM-Autoencoders höhere Genauigkeiten und eine bessere Erkennung komplexer oder unbekannter Angriffsmuster erzielen. Unterschiede in der Erkennungsleistung ergeben sich insbesondere in Abhängigkeit vom Angriffstyp.

Die Ergebnisse verdeutlichen, dass KI-gestützte Verfahren ein erhebliches Potenzial für die IT-Sicherheit besitzen, ihre Praxistauglichkeit jedoch stark von der Qualität der Datensätze, den benötigten Ressourcen und der Transparenz der Modelle abhängt. Zukünftige Entwicklungen wie Explainable AI (XAI), Federated Learning oder Large Language Models (LLM) bieten vielversprechende Ansätze, um bestehende Grenzen zu überwinden.

Abstract

This thesis addresses the detection of Distributed Denial of Service (DDoS) attacks using Artificial Intelligence (AI) methods. With the increasing frequency and complexity of such attacks, traditional defence mechanisms such as firewalls and intrusion detection systems are reaching their limits. The aim of this work is to analyse various Machine-Learning (ML) and Deep-Learning (DL) approaches and to compare their effectiveness in detecting different attack types.

The theoretical part introduces the fundamentals of DDoS attacks, their type and impacts, as well as the functioning of ML and DL methods. A comparative analysis of model performances based on study results follows. The findings show that ML models such as Random Forest or SVM are efficient and easy to implement, while DL models such as CNNs or LSTM-Autoencoders achieve higher accuracies and are more capable of identifying complex or previously unknown attack patterns. Detection performance strongly depends on the attack type.

The results highlight that AI-based approaches hold great potential for IT security, but their practical applicability depends heavily on dataset quality, resource requirements and model transparency. Future developments such as Explainable AI (XAI), Federated Learning or Large Language Models (LLM) provide promising directions to overcome current limitations.

Inhaltsverzeichnis

Abkürzungsverzeichnis

AE	Autoencoder
AGI	Artificial General Intelligence
AUC	Area Under the ROC Curve
CIA	Confidentiality, Integrity, Availability
CIAC	Computer Incident Advisory Capability
CNN	Convolutional Neural Network
CR	Klassifizierungsrate
DIN	Deutsches Institut für Normung
DDoS	Distributed Denial of Service
DNS	Domain Name System
DoF	Denial of Firewalling
DoS	Denial-of-Service
FL-DAD	Federated Learning
FN	False Negative
FP	False Positive
FMC	F-Measure Complement
FNR	False Negative Rate
FPR	False Positive Rate
GB	Gradient Boosting
GenAI	Generative Artificial Intelligence
HPC	Hardware Performance Counter
IDPS	Intrusion Detection Prevention System
IDS	Intrusion Detection System
IoT	Internet of Things
KI	Künstliche Intelligenz
KNN	K-Nearest-Neighbor
LLM	Large Language Model

LSTM	Long Short-Term Memory
MADRL	Multi-Agent Deep Reinforcement Learning
MCC	Mobile Cloud Computing
MDP	Markov-Entscheidungsprozess
ML	Machine-Learning
MTT	Mitigation Thoughput Time
NPV	Negative Predictive Value
NTP	Network Time Protocol
ResNet	Residual Network
SDN	Software Defined Network
SOM	Self-Organized-Map
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TTL	Time to Live
TP	True Positive
XAI	Explainable Artifical Intelligence

Abbildungsverzeichnis

1 Einleitung und Aufbau

Die Digitalisierung nimmt in der heutigen Zeit drastisch und kontinuierlich zu. Parallel dazu nehmen auch die Cyberangriffe deutlich in ihrer Anzahl sowie ihrer Komplexität zu. So sind stabile und sichere Netzwerke von großer Bedeutung. In dieser Arbeit wird auf einen der am häufigsten angewendeten Cyberangriffe eingegangen, und zwar das Distributed Denial of Service (DDoS). Dieser beeinträchtigt stark die Verfügbarkeit von Diensten durch die gezielte Überlastung.

Diese Arbeit befasst sich mit der Untersuchung von den verschiedenen Methoden zur Erkennung der DDoS-Angriffe. Der Fokus liegt hierbei auf dem Einsatz von Künstlicher Intelligenz (KI), insbesondere Verfahren des Machine Learning (ML) und Deep Learning (DL). Um auch komplexe und variierende Angriffsmuster zuverlässig zu erkennen, stellen hierbei die KI-basierten Verfahren eine vielversprechende Möglichkeit dar. In der folgenden Arbeit werden daher unterschiedliche KI-gestützte Ansätze vorgestellt, analysiert und hinsichtlich ihrer Wirksamkeit bewertet.

Aufbau der Arbeit:

In Kapitel 2 werden die Grundlagen von DDoS-Angriffen erläutert, typische Angriffstypen vorgestellt und klassische Schutzmechanismen beschrieben. Zudem werden die Konzepte von ML und DL sowie deren Relevanz für die Angriffserkennung dargestellt. Kapitel 3 beschreibt die Funktionsweise von KI-gestützten Erkennungssystemen, definiert Anforderungen und Evaluationsmetriken und stellt verschiedene Modelle im Detail vor. In Kapitel 4 folgt eine vergleichende Analyse der Leistungsfähigkeit dieser Modelle im Hinblick auf unterschiedliche Angriffstypen, ergänzt durch eine kritische Diskussion ihrer Grenzen und Potenziale. Kapitel 5 widmet sich aktuellen Trends und Weiterentwicklungen in der KI-gestützten Cybersicherheit. Kapitel 6 fasst die zentralen Ergebnisse zusammen, beantwortet die Forschungsfrage und gibt einen Ausblick auf zukünftige Forschungs- und Anwendungsperspektiven.

1.1 Problemstellung

Da DDoS-Angriffe gezielt auf die Verfügbarkeit von Diensten abzielen, stellen sie ein großes Sicherheitsrisiko für Netzwerke dar. Zudem stellt auch ihre Vielseitigkeit eine große Bedrohung dar. Durch ihre volumetrischen Angriffe können Angreifer Netzwerke durch riesige Datenfluten überlasten. Protokollbasierte oder anwendungsorientierte Angriffe zielen direkt auf Schwachstellen in Netzwerken ab, was sie somit schwerer erkennbar macht.

Es ist notwendig, neue Erkennungsmethoden zu entwickeln, da traditionelle Schutzmechanismen wie Firewalls oder Intrusion Detection Systeme (IDS) bei solchen Angriffen schnell an ihre Grenzen stoßen. KI-gestützte Systeme stellen hier eine große Rolle dar, da sie in der Lage sind, komplexe Muster im Netzwerkverkehr zu analysieren. Die Herausforderung dabei ist es jedoch, zu bestimmen, wie zuverlässig diese Systeme die unterschiedlichen Angriffstypen erkennen können und welche Unterschiede in ihrer Leistungsfähigkeit bestehen.

1.2 Zielsetzung und Forschungsfrage

Das Ziel dieser Arbeit ist es, herauszufinden, wie gut verschiedene KI-Systeme darin sind, verschiedene Arten von DDoS-Angriffen zu erkennen. Diese werden dann analysiert und miteinander verglichen, wobei Unterschiede in Genauigkeit, Fehlklassifikationen und Robustheit der Modelle herausgearbeitet werden. Auf Grundlage dieser Ergebnisse werden Rückschlüsse auf die Effektivität der jeweiligen Ansätze sowie deren Unterschiede gezogen.

Forschungsfrage:

„Wie effektiv erkennen KI-gestützte Systeme verschiedene DDoS-Angriffstypen und welche Unterschiede bestehen in ihrer Erkennungsleistung?“

2 Grundlagen

In diesem Kapitel werden die theoretischen Grundlagen, die für das Verständnis dieser Arbeit erforderlich sind, vermittelt. Der Fokus liegt hierbei darauf, zentrale Begriffe und Konzepte einzuordnen und ein gemeinsames Verständnis zu schaffen, das für die spätere Analyse und Diskussion notwendig ist.

2.1 DDoS-Angriffe: Definition, Ziele und Auswirkungen

Als einer der größten und bekanntesten Bedrohungen in der Cybersicherheit gehören die DDoS-Angriffe. Deren Ziel ist es, mit Flooding Angriffe, den Zugriff von legitimen Nutzern auf einen Dienst zu stören oder komplett zu verhindern. Bei den Flooding Angriffen, werden die Infrastrukturen des Opfers mit riesigen Anzahlen von Anfragen überlastet bzw. „überflutet“. Dies wird von den Angreifern erreicht, indem sie die Schwachstellen in den Computersystemen ausnutzen. Somit können sie eine Vielzahl kompromittierter Geräte, sogenannte Botnets, aufbauen. Sobald dieses Botnets Netzwerk aufgebaut und aktiv ist, kann der Angreifer einen groß angelegten Angriff gegen ausgewählte Zielsysteme starten (Zargar et al., 2013).

Der erste dokumentierte Vorfall eines DDoS-Angriffes ist der Angriff auf die US-Amerikanische Computer Incident Advisory Capability (CIAC) im Jahr 1999. Jedoch sind die DoS-Angriffe seit den 1980er Jahren bereits bekannt (Zargar et al., 2013).

DDoS-Angriffe lassen sich in zwei Hauptkategorien einteilen. Einmal den Schwachstellenangriff. Hierbei handelt es sich um fehlerhaft aufgebaute Datenpakete, die ein Angreifer senden kann, um Protokolle oder Anwendungen gezielt zu verwirren. Jedoch sind Flooding-Angriffe viel häufiger. Hier werden massive Mengen an Datenverkehr erzeugt, um entweder Netzwerkressourcen (wie Bandbreite oder Rechenkapazitäten von Routern) oder direkt die Serverressourcen (wie CPU, Speicher oder offene Verbindungen) zu überlasten (Zargar et al., 2013).

Es ist zudem auch wichtig, zwischen einem klassischen Denial of Service-Angriff (DoS) und einem DDoS-Angriff zu unterscheiden. Ein DDoS-Angriff wird durch mehrere verteilten Systeme gleichzeitig

verursacht, während bei einem DoS-Angriff ein einzelner Angreifer versucht, ein System mit Anfragen zu überfluten. Da nur ein einzelner Angreifer beteiligt ist, kann die Quelle eines DoS-Angriffs leichter nachvollzogen werden. Bei einem DDoS-Angriff fällt dies deutlich schwerer, aufgrund der Vielzahl an Angriffsquellen. Außerdem ist die Wirkung eines DDoS-Angriffes meist deutlich stärker, da hier mehrere kompromittierte Systeme gleichzeitig agieren (Siriyapuraju et al., 2023).

Aktuelle DDoS-Angriffe werden fast ausschließlich durch ferngesteuerte Botnets oder sogenannte Zombies durchgeführt. Diese Geräte senden gleichzeitig große Datenmengen oder Anfragen an das Zielsystem. Die Folge dabei ist, dass das System extrem langsam reagiert oder im schlimmsten Fall vollständig abstürzt. Eine zusätzliche Herausforderung für Verteidigungsmaßnahmen besteht darin, dass Angreifer oft gefälschte IP-Adressen verwenden, was die Identifizierung erheblich erschwert (Zargar et al., 2013).

Viele Angreifer haben das Hauptziel, mit den DDoS-Angriffen Dienste lahmzulegen. Diese haben meistens enorme wirtschaftliche Folgen für die Betroffenen. Neben direkten Umsatzeinbußen entstehen auch Kosten für Abwehrmaßnahmen und Wiederherstellungsprozesse. Es könnte aber auch zu Imageverlusten kommen sowie zu Dienstunterbrechungen und langfristigen finanziellen Schäden. Dabei werden Techniken eingesetzt wie die bloße Überflutung mit Datenmengen bis hin zur gezielten Manipulation des Netzwerkverkehrs oder dem Ausnutzen konkreter Schwachstellen zur Erlangung unbefugten Zugriffs (Zargar et al., 2013).

Die Auswirkungen solcher Angriffe sind sehr gravierend. Dies wurde besonders deutlich 2022, als Cloudflare einen Anstieg von DDoS-Angriffen um 60% im Vergleich zum Vorjahr meldete. Dabei waren insbesondere cloudbasierte Anwendungen und Infrastrukturen betroffen (Ali, 2025).

Die Cybersicherheitsmaßnahmen müssen kontinuierlich weiterentwickelt werden, denn mit dem technologischen Fortschritt werden diese Bedrohungen genauso ausgebaut. Somit handelt es sich dabei um einen regelrechten Wettlauf zwischen Angreifern und Verteidigern (Zargar et al., 2013).

Es gibt verschiedene Motive für DDoS-Angriffe, zum einen das finanzielle Interesse, wie die Erpressung oder das Lahmlegen der Konkurrenz, oder Racheakte sowie politische Beweggründe. Hierzu gibt es ein aktuelles Beispiel, welches sich im Mai 2025 abgespielt hat. Die pro-russische Hackergruppe „NoName057“ bekannte sich zu einer Reihe koordinierter Angriffe auf britische Websites, darunter Regierungsportale und lokale Behörden. Ihr Ziel war es dabei, die betroffenen Seiten durch massiven Datenverkehr lahmzulegen. Einige dieser Seiten waren für eine kurze Zeit nicht mehr erreichbar, jedoch konnten sie nach wenigen Stunden wiederhergestellt werden. Die Erklärung der Gruppe NoName057 dabei war, dass Großbritannien am Ukrainekonflikt beteiligt wäre. Die Gruppe ist auch bereits seit 2022 durch ähnliche Attacken auf europäische und amerikanische Regierungsseiten aufgefallen (Boffey, 2025).

2.2 Typen von DDoS-Angriffen

DDoS-Angriffe lassen sich in drei Hauptkategorien einteilen: volumetrische Angriffe, protokollbasierte Angriffe und Application-Layer-Angriffe. Im Folgenden werden diese Kategorien und ihre typischen Angriffsformen näher erläutert.

Volumetrische Angriffe zielen darauf ab, die Bandbreite oder Kapazitäten eines Netzwerkes zu überlasten. Dies geschieht entweder durch das massenhafte Versenden von Anfragen, um die Serviceverarbeitung auszulasten, oder durch das Vergrößern der Datenpakete, um die Netzwerkressourcen wie die Bandbreite zu erschöpfen (Nogueria et al., 2017).

Ein klassisches Beispiel ist der DNS-Amplification-Angriff, der in Abbildung 1 veranschaulicht wird. Hierbei wird das Domain Name System (DNS) ausgenutzt, das Domainnamen in IP-Adressen übersetzt. Angreifende nutzen offene DNS-Resolver und senden gefälschte DNS-Anfragen, bei denen die IP-Adresse des Opfers als Absender angegeben ist. Die Antwortpakete der DNS-Server, die deutlich größer als die ursprünglichen Anfragen sind, werden an das Opfer gesendet, wodurch der Datenverkehr verstärkt und das Zielnetzwerk überlastet und gleichzeitig unerreichbar wird (Mekala et al., 2024).

Abbildung 1: DNS-Amplification-Angriff. Eigene Darstellung

Ein weiterer Angriff dieser Art ist der UDP-Flood-Angriff. In der folgenden Abbildung 2 von Singh und Juneja (2010, S. 3406) wird dies veranschaulicht.

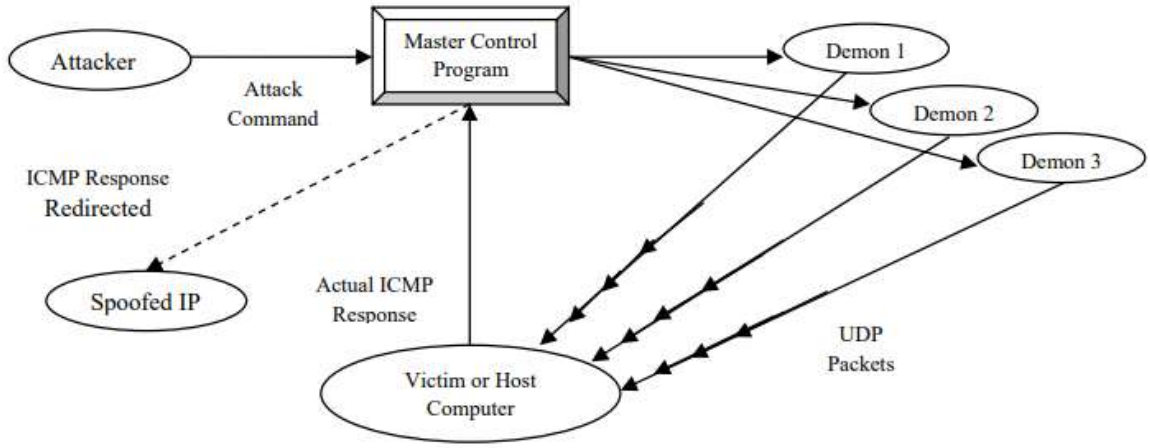

Abbildung 2: UDP-Flood Attack Mechanism (Quelle: Singh und Juneja, 2010, S.3406)

In der Abbildung 2 sieht man deutlich, wie der angreifende Host einen Befehl an ein Master Control Programm sendet, welches als Angriffs-Handler dient. Somit instruiert das Master Control Programm wiederum Agenten wie Demons oder Zombies (kompromittierte Systeme), in diesem Fall Demons.

Diese Demons verschicken daraufhin massenhaft UDP-Pakete mit gefälschter Absenderadresse an das Opfer. Das System versucht, auf diese Pakete zu antworten, was jedoch ins Leere läuft und letztendlich zur Erschöpfung der Ressourcen und zum Absturz führen kann. Außerdem kann jedes Master-Control-Programm mehrere Demons steuern und auch können mehrere parallel aktiv sein. Dies sorgt für eine massive Anzahl von UDP-Paketen, die gleichzeitig an das Zielsystem gesendet werden. Dies führt dazu, dass das Zielsystem sogar schneller überflutet wird (Singh und Juneja, 2010).

Eine zusätzliche Methode besteht darin, UDP-Pakete an zufälligen, möglicherweise geschlossenen Ports zu senden. Das Opfer antwortet mit ICMP-Paketen („Port unreachable“), was bei einer großen Menge zur Überlastung führt (Treseangrat et al., 2012).

Ein verwandter Angriff ist die NTP-Amplification, die das Network Time Protocol (NTP) missbraucht. Hierbei senden Angreifer kleine Anfragen mit gefälschter IP-Adresse (die des Opfers) an die NTP-Server, die durch den aktivierten MONLIST-Befehl umfangreiche Antworten zurücksenden, denn wenn dieser Befehl genutzt wird, gibt der Server eine Liste mit den letzten 600 IP-Adressen zurück, die ihn kontaktiert haben. Diese „Verstärkung“ führt dazu, dass eine Vielzahl von NTP-Servern gleichzeitig riesige Datenmengen an das Opfer überträgt, was dessen Netzwerkanschluss überflutet. Hierbei bleibt der Angreifer auch unsichtbar, da er seine eigene IP-Adresse nie verwendet (Rudman und Irwin, 2015).

Protokollbasierte Angriffe zielen auf Schwachstellen in Netzwerkprotokollen ab. Ein typisches Beispiel ist der SYN-Flood-Angriff, der den TCP-Handshake ausnutzt.

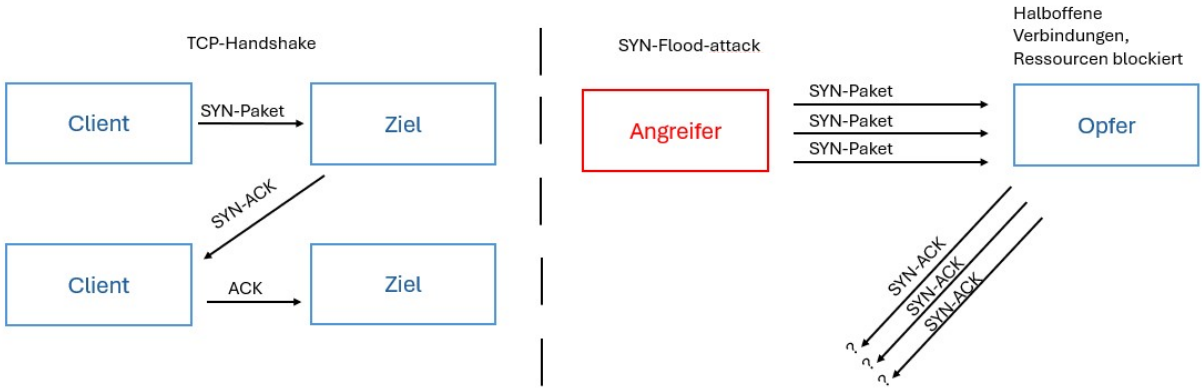

Abbildung 3: TCP-Handshake vs SYN-Flood-attack. Eigene Darstellung

Wie man anhand von Abbildung 3 sehen kann, besteht der Verbindungsaufbau normalerweise aus einem dreistufigen Prozess: Der Client sendet ein SYN-Paket, der Server antwortet mit einem SYN/ACK-Paket und der Client bestätigt mit einem ACK-Paket. Beim SYN-Flood sendet der Angreifer jedoch nur das erste Paket und unterlässt die Bestätigung. Der Server wartet auf die Vervollständigung der Verbindung und hält dabei Ressourcen für jede unvollständige Anfrage bereit. Bei vielen solcher halb offenen Verbindungen kommt es zur Erschöpfung der Verbindungstabellen oder des Arbeitsspeichers, wodurch legitime Anfragen nicht mehr bearbeitet werden können. Die Auswirkungen eines SYN-Flood-Angriffs auf die CPU-Auslastung werden deutlich im folgenden Zitat beschrieben: „In a normal scenario, CPU

usage is about 20%. However in attacked phases with the same CPU load, CPU execution overhead is nearly 90% or 100%.“ (Kukreti et al., 2022, S. 323).

Application-Layer-Angriffe, auch Layer-7-DDoS genannt, zielen auf die Anwendungsschicht des OSI-Modells, in der unter anderem Webanwendungen verarbeitet werden. Sie sind besonders tückisch, da sie den legitimen Datenverkehr imitieren und schwer zu erkennen sind.

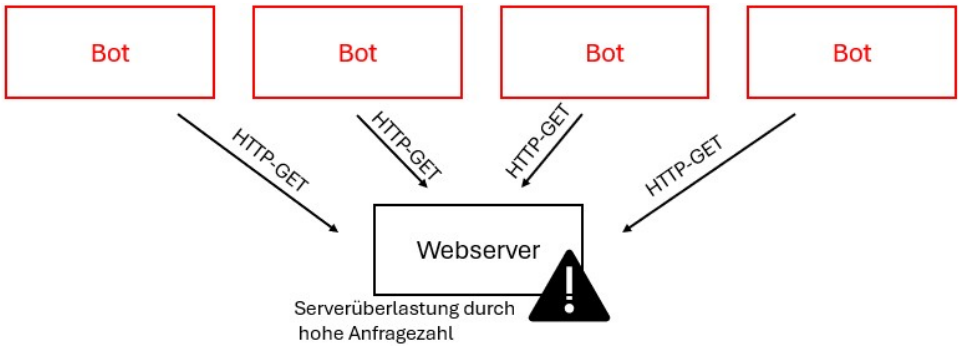

Abbildung 4: HTTP-Flood. Eigene Darstellung

Wie in Abbildung 4 zu sehen ist, werden beim HTTP-Flood-Angriff zahlreiche HTTP-GET- oder POST-Anfragen an einen Webserver gesendet. Diese wirken authentisch, zwingen den Server jedoch, jede Anfrage zu verarbeiten, inklusive möglicher Datenbankabfragen oder API-Aufrufe, was Rechenleistung, Speicher und andere Ressourcen verbraucht. Die Ressourcen des Servers werden dadurch überlastet, was in einer „503 Server Unavailable“-Fehlermeldung resultieren kann. Angreifer verwenden häufig Botnets, um die Anfragen von verschiedenen IP-Adressen aus zu versenden und die Erkennung zu erschweren (Varre und Bayana, 2022).

Außerdem gibt es noch den Low-Rate-Angriff, welcher auch überwiegend in Layer 7 stattfindet. Dies ist ein langsamer und unauffälliger Angriff. Statt mit vielen Daten auf einmal zu überfluten, schickt der Angreifer über längere Zeit kleine Datenmengen. Dabei nutzt er die begrenzte Speicherkapazität in den Switches von Software Defined Network (SDN)-Netzwerken, einer Architektur, bei der die Steuerungsebene von der Datenebene getrennt ist und zentrale Controller die Regeln für den Datenverkehr verwalten. Dadurch werden immer mehr Regeln eingetragen, bis die Tabelle voll ist. Wenn das passiert, können legitime Anfragen nicht mehr verarbeitet werden und es kommt zu einer Dienstverweigerung (Zhijun et al., 2020).

Eine spezielle Form dieses Angriffs ist der Slowloris-Angriff, bei dem Verbindungen über lange Zeit offengehalten werden. Der Angreifer sendet eine Vielzahl unvollständiger HTTP-Anfragen, bei denen die Header-Informationen nur sehr langsam übermittelt werden. Der Webserver hält diese Verbindungen offen, was dazu führt, dass legitime Anfragen nicht mehr bearbeitet werden können. Slowloris ist besonders effektiv gegen Webserver mit thread-basiertem Verbindungsmanagement (z.B. Apache) und schwer zu erkennen, da die Bandbreitenauslastung gering bleibt (Oktivasari et al., 2022).

Zusammenfassend zeigen die verschiedenen Typen von DDoS-Angriffen, wie vielseitig und zugleich gefährlich diese Angriffsmethoden sein können. Während volumetrische Angriffe vor allem die Bandbreite lahmlegen, zielen Protokoll- und Application-Layer-Angriffe gezielter auf spezifische Schwachstellen in Systemen. Besonders gefährlich sind solche Angriffe, die legitimen Datenverkehr imitieren, da sie schwer zu identifizieren sind. Die Kenntnis der einzelnen Angriffsarten sind somit essenziell für die Entwicklung effektiver Abwehrmechanismen.

Ein besonderer Fokus liegt in der aktuellen Forschung und Bedrohungslage auf Angriffen, die durch Internet of Things (IoT) ermöglicht werden. Hierbei werden moderne und äußerst relevante Formen von DDoS-Angriffen dargestellt, da diese auf neuartige Infrastrukturen mit massenhaft vernetzten Geräten abzielen. Im folgenden Abschnitt wird diese Angriffskategorie näher betrachtet.

2.2.1 IoT-basierte DdoS-Angriffe

Um einen automatisierten Datenaustausch und eine Fernsteuerung zu ermöglichen, verbindet IoT alltägliche physische Geräte wie Überwachungskameras, Smart-Home-Anwendungen oder Industrieanlagen mit dem Internet. Dadurch, dass diese Geräte untereinander Informationen austauschen, ermöglich diese, sie bei Bedarf zu steuern und zu nutzen. In den letzten Jahren hat IoT ein enormes Wachstum erlebt, was zu einem unendlichen und kontinuierlichen Anstieg der erzeugten und übertragenen Datenmengen geführt hat (Srivastava et al., 2024).

Die digitale und physische Welt wird zunehmend miteinander verbunden, da das IoT heute in zahlreichen Bereichen Anwendung findet. In der Landwirtschaft, im Gesundheitswesen oder in der Industrie. Diese zunehmende Vernetzung bringt jedoch auch große Sicherheitsrisiken mit sich. Besonders im Hinblick auf DDoS-Angriffe zeigt sich, dass viele IoT-Geräte nur über begrenzte Verarbeitungsleistung verfügen und häufig mit unzureichenden Sicherheitsmaßnahmen ausgestattet sind. So werden diese Schwächen gezielt von Angreifern ausgenutzt, um die Kontrolle über Geräte zu erlangen. Die infizierten Geräte bilden dann die Botnetze, die gemeinsam koordinierte Angriffe auf leistungsfähigere Zielinfrastrukturen ausführen (Ashfaq et al., 2022; Srivastava et al., 2024).

Das Hauptziel dieser Angriffe ist es, die Bandbreite eines Servers mit massenhaftem Datenverkehr zu überlasten, um den Dienst für legitime Nutzer unzugänglich zu machen. Durch die weltweit rasant wachsende Zahl an IoT-Geräten vergrößert sich die potenzielle Angriffsfläche rapide (Srivastava et al., 2024).

Ein bekanntes Beispiel für ein solches IoT-Botnet ist Mirai, das 2016 weltweit Aufmerksamkeit erregte. Es kompromittierte hunderttausende ungenügend gesicherte IoT-Geräte wie Kameras und Router und nutzte sie für massive DDoS-Angriffe, die unter anderem den DNS-Anbieter Dyn zeitweise lahmlegten. Dieser Vorfall verdeutlichte eindrücklich, wie groß die Bedrohung durch IoT-Infrastrukturen mit geringen Sicherheitsstandards sind und welche Folgen Botnetze im globalen Maßstab haben können (Kolias et al., 2017).

IoT-basierte DDoS-Angriffe zählen somit zu den spezifischen Formen von Cyberangriffen, die gezielt auf internetfähige Geräte abzielen. IoT-Umgebungen sind besonders anfällig dafür, Teil riesiger Botnetze zu werden, aufgrund der Kombination aus begrenzter Rechenkapazität, mangelhaften Sicherheitsvorkehrungen und hoher Verarbeitung (Ashfaq et al., 2022; Srivastava et al., 2024).

2.3 Grundlagen der IT-Sicherheit im Kontext von DDoS

In den vorherigen Kapiteln wurde verdeutlicht, wie komplex und unterschiedlich DDoS-Angriffe anhand der verschiedenen Angriffsarten sein können. Zum einen durch die massive Bandbreitenüberlastung und zum anderen durch die schwer erkennbaren Angriffe auf die Anwendungsebenen. Dadurch, dass die Bedrohungslage kontinuierlich weiter zunimmt, werden umfassende Schutzmaßnahmen benötigt, um diese wirksam entgegenzustehen.

In den folgenden Unterkapiteln werden zunächst die relevanten Schutzziele im Umgang mit DDoS-Angriffen erläutert. Darauf aufbauend werden die konkreten Schutzmaßnahmen vorgestellt. Abschließend wird die Bedeutung von Monitoring und Reaktion im Ernstfall erläutert.

2.3.1 Relevante Schutzziele

DDoS-Angriffe stellen eine erhebliche Bedrohung für die Informationssicherheit dar, da sie primär auf die Verfügbarkeit von Systemressourcen abzielen. Dabei kann die Verfügbarkeit, welche eines der drei zentralen Schutzziele der Informationssicherheit im Rahmen der sogenannten CIA-Trade (Confidentiality, Integrity, Availability) darstellt, gezielt kompromittiert werden (Pathak et al., 2025). Die CIA-Trade bildet ein grundlegendes Modell zur Steuerung und Bewertung der Informationssicherheit innerhalb einer Organisation. Bereits die Beeinträchtigung eines dieser drei Kernprinzipien kann zu einem erheblichen Risiko führen (Raja et al., 2022).

Die Verfügbarkeit stellt dabei das Hauptziel von DDoS-Angriffen dar. Angriffe dieser Art können Dienste für ihre eigentlichen Nutzer unzugänglich machen oder Systemressourcen vollständig zerstören, was sie laut Raja et al. (2022, S. 1) zu einer der „most disastrous attack [sic]“ macht.

Neben der direkten Gefährdung der Verfügbarkeit können DDoS-Angriffe auch indirekte Auswirkungen auf die Vertraulichkeit und Integrität von Informationen haben. So werden sie laut Pathek et al. (2025) mitunter als Ablenkungsmanöver eingesetzt, um parallel sensible Daten zu stehlen. In solchen Fällen wird die Vertraulichkeit verletzt, da private Informationen unberechtigt erlangt werden. Auch die Integrität von Informationen kann in bestimmten Szenarien gefährdet sein. Angreifer könnten durch den Zugriff auf Cloud-Ressourcen Veränderungen vertraulicher Daten vornehmen, was zu einer Kompromittierung der Datenintegrität führen würde (Pathek et al., 2025; Raja et al., 2022).

Zusammenfassend lässt sich feststellen, dass DDoS-Angriffe zwar primär die Verfügbarkeit bedrohen, jedoch auch potenzielle, wenn auch indirekte, Risiken für die Vertraulichkeit und Integrität darstellen.

2.3.2 Schutzmaßnahmen gegen DDoS-Angriffe

Es gibt vielfältige klassische Schutzmaßnahmen gegen DDoS-Angriffe, die unterschiedliche Angriffstypen abwehren sollen. Eine dieser Schutzmaßnahmen nennt sich Rate Limiting. Sie dient als Schutz gegen flooding-basierte DDoS-Angriffe. Dieses Verteidigungsframework besteht aus drei Hauptkomponenten: Erkennung, IP-Rückverfolgung und Bandbreitenkontrolle. Es wird auf allen Edge-Routern des Netzwerks eingesetzt und bietet Schutz gegen DDoS-Angriffe durch Kommunikation zwischen dem Edge-Router am Quellende und dem Edge-Router am Opferende.

Die erste Hauptkomponente ist die distanzbasierte DDoS-Erkennungskomponente. Sie implementiert einen existierenden distanzbasierten DDoS-Erkennungsalgorithmus, bei dem die Entfernung eines Pakets durch das TTL-Feld (Time to Live) im IP-Header berechnet wird. Der IP-Header enthält Steuerinformationen eines IP-Pakets, wie Quell- und Zieladresse sowie das TTL-Feld. Das TTL-Feld gibt an, wie viele Router das Paket maximal passieren darf, bevor es verworfen wird, um eine Endlosschleife im Netzwerk zu verhindern. Der Entfernungswert am Opferende wird berechnet, indem man den endgültigen TTL-Wert vom Startwert abzieht. Die DDoS-Erkennungskomponente wird im Edge-Router am Opferende eingesetzt. Der Router erkennt anomale Veränderungen im eingehenden Traffic und trennt Angriffstraffic von legitimen Anfragen. Dabei werden das Konzept der mittleren absoluten Abweichung und die exponentielle Glättungstechnik verwendet, um den durchschnittlichen Entfernungswert und die Verkehrsrate für das nächste Zeitintervall vorherzusagen. Die mittlere absolute Abweichung ist ein Maß dafür, wie stark einzelne Werte im Durchschnitt vom Mittelwert abweichen. Die exponentielle Glättungstechnik berücksichtigt frühere Werte mit abnehmendem Gewicht zur Prognose zukünftiger Werte. Der tatsächliche Wert wird mit dem vorhergesagten verglichen und eine Anomalie-Flag wird gesetzt, wenn der tatsächliche Wert außerhalb des vom Vorhersagewert definierten Bereichs liegt.

Die zweite Komponente ist das Bandbreitenkontrollmodul. Diese nutzt das Konzept der Ratenbegrenzung, um Angriffstraffic am Quellende zu verwerfen. Dabei werden die Angreiferknoten nicht alle gleich behandelt, sondern je nach Aggressivität des gesendeten Traffics und Kapazität des Kanals am Opferende unterschiedlich berücksichtigt. Die Berechnung des Rate Limits basiert auf diesen beiden Faktoren. Obwohl diese Technik effektiv gegen massenhaften Traffic sein kann, stößt sie bei verteilten Botnet-Angriffen oder Low-Rate-Angriffen an ihre Grenzen (Patil und Ragha, 2011).

Neben dem Rate Limiting stellt auch der Einsatz von Firewalls eine etablierte Schutzmaßnahme dar. Firewalls sind dazu entwickelt, ein System vor unautorisiertem Zugriff zu schützen und eine Barriere zwischen privaten und öffentlichen Netzwerken zu bilden. Sie blockieren unerwünschten Datenverkehr und lassen nur autorisierten Datenverkehr zu. Somit verhindern sie im Falle eines DDoS-Angriffs, dass bösartiger Datenverkehr auf andere Server oder Systeme übergreift (Teja et al., 2023).

Ein Beispiel hierfür sind die Zugriffskontrolllisten (Access Control Lists, ACLs), die von Administratoren innerhalb der Firewalls konfiguriert werden, um autorisierten IP-Adressen, Domain-Namen, Protokollen, Programmen, Portnummern und Schlüsselwörtern den Zugriff auf den Server oder das System zu ermöglichen und unautorisierten Zugriff zu verweigern (Teja et al., 2023).

Zusätzlich bieten zustandsbehaftete Firewalls (Stateful Firewalls) ein höheres Sicherheitsniveau als einfache Paketfilter. Hierbei wird eine Statustabelle verwendet, um aktuelle Pakete mit früheren abzugleichen. Dabei werden Sicherheitslücken geschlossen, die von Angreifern für DoS-Angriffe sonst ausgenutzt werden könnten. Die Firewalls können alle Pakete verfolgen und deren Ziel bestimmen, indem sie in ihren Sitzungstabellen, die Aufzeichnungen über vergangene Verkehrsflüsse speichern. Auch bei einer hohen Anzahl an Anfragen, können die Tabellen um Einträge erweitert werden, um die Sicherheit der Firewall zu verbessern und unerwünschten Datenverkehr zu blockieren (Teja et al., 2023).

Obwohl Firewalls einen erheblichen Schutz bieten, sind sie dennoch anfällig für verschiedene Arten von Angriffen, einschließlich DDoS-Angriffen. Sie müssen zusätzliche Anstrengungen bei einem Angriff übernehmen, um den Datenverkehr zu verhindern, was jedoch ihre Leistung sehr beeinträchtigen kann. Die Sitzungstabellen können bei gezielten Angriffen mit falschen Einträgen überflutet werden, was dann dazu führt, dass legitimer Datenverkehr blockiert wird. Solche Angriffe werden als Denial of Firewalling (DoF) bezeichnet. Firewalls können durch DoF-Angriffe unempfindlich gemacht werden, womit sie dann für eine längere Zeit ihre Fähigkeiten als Schutzbarriere verlieren. DoF nutzt Schwachstellen von Firewalls aus und überflutet die Sitzungstabellen von Firewalls mit falschen oder massenhaften Einträgen. Normalerweise verbessern Sitzungstabellen die Firewall Leistung, indem sie wiederkehrenden legitimen Datenverkehr, der bereits überprüft wurde, nicht erneut analysieren müssen. Wenn jedoch diese Tabellen durch Angreifer mit fehlerhaften Daten gefüllt werden, führt dies dazu, dass legitime Verbindungen nicht mehr hergestellt werden können und die Firewall ihre unterstützende Funktion verliert (Teja et al., 2023).

Eine weitere wichtige Schutzmaßnahme gegen DDoS-Angriffe stellt das Intrusion Detection and Prevention System (IDPS) dar. Vorab wird das Intrusion Detection System (IDS) erläutert.

Ein IDS ist ein Sicherheitssystem, das den Netzwerkverkehr oder die Systemaktivitäten überwacht, um Anzeichen für potenzielle Angriffe oder Richtlinienverstöße zu erkennen. Es funktioniert ähnlich wie ein Alarmsystem: Wird ein verdächtiges Muster erkannt, wird eine Warnung ausgegeben, sodass Administratoren reagieren können (Mazhar et al., 2021).

Es gibt einerseits die signaturbasierten IDS-Systeme. Viele klassische IDS-Systeme verwenden Musterabgleichs- und signaturbasierte Ansätze. Dies heißt, dass sie nach bekannten Mustern von Angriffen im Datenverkehr suchen (Signaturen). Jedoch kommt dies mit einer Einschränkung, denn neue Angriffsarten werden hierbei nicht erkannt, da von ihnen keine Signaturen vorhanden sind. Andererseits gibt es anomaliebasierte IDS-Systeme. Hier verwenden IDS-Systeme statische Techniken oder maschinelles

Lernen, um Anomalien (Abweichungen vom normalen Verhalten) zu erkennen. Dies stellt sich als effektiver dar, um neue Angriffsarten zu entdecken. Das Problem hierbei ist jedoch, dass es zu einer hohen Fehlalarmrate kommen kann (False Positives) (Mazhar et al., 2021).

IDS sind nicht für den Einsatz auf IoT-Geräten geeignet, da diese Geräte oft über begrenzte Ressourcen in Bezug auf Speicher, Rechenleistungen und Energieverbrauch verfügen. An dieser Stelle kommt das IDPS zum Einsatz. IDPS ist eine Weiterentwicklung des IDS. Sie erkennt nicht nur Eindringversuche und bedrohliche Aktivitäten, sondern kann diese auch automatisch verhindern oder abwehren. Ein wesentliches Problem hierbei ist der hohe Ressourcenverbrauch, welcher die Leistung des Netzwerkes beeinträchtigen kann (Mazhar et al., 2021).

Eine weitere zentrale Komponente klassischer Schutzmaßnahmen ist Monitoring und Reaktion. Monitoring-Systeme überwachen den Netzwerkverkehr in Echtzeit, analysieren Auffälligkeiten und lösen Alarme bei verdächtigen Aktivitäten aus. Sie können sowohl eigenständig als auch in Kombination mit Firewalls oder IDPS-Systemen eingesetzt werden. Effektive Monitoring-Ansätze erlauben nicht nur die Identifikation von Angriffen, sondern auch eine schnelle Reaktion, etwa durch automatisches Filtern auffälliger Datenströme oder die Umleitung von Traffic. Hierdurch lässt sich die Auswirkung eines DDoS-Angriffs deutlich reduzieren, indem die Erkennung und Reaktion enger miteinander verzahnt werden. Allerdings stoßen Monitoring- und Reaktionssysteme in der Praxis auch an ihre Grenzen. Sie erfordern oft hohe Rechenressourcen, können Fehlalarme erzeugen und sind bei sehr groß angelegten Angriffen selbst einem Überlastungsrisiko ausgesetzt (Padhiar et al., 2025).

Im Kontext des CIA-Triads wird deutlich, dass DDoS-Angriffe in erster Linie die Verfügbarkeit bedrohen, da Dienste für legitime Nutzer nicht mehr erreichbar sind. Klassische Schutzmaßnahmen dienen daher vor allem der Aufrechterhaltung der Verfügbarkeit. Indirekt tragen sie jedoch auch zur Integrität bei, indem sie verhindern, dass Angriffe zu fehlerhaften Zuständen oder Systemabstürzen führen. Die Vertraulichkeit bleibt bei klassischen DDoS-Angriffen zwar unberührt, kann aber gefährdet werden, wenn Angriffe als Ablenkungsmanöver für Datenexfiltration genutzt werden.

Die genannten Einschränkungen machen deutlich, dass klassische Schutzmaßnahmen allein nicht ausreichen, um moderne DDoS-Angriffe zuverlässig abzuwehren. Insbesondere verteilte Botnets, neuartige Angriffsmuster und die Skalierung im IoT-Umfeld erfordern flexiblere Ansätze (Zargar et al., 2023). An dieser Stelle setzen KI-gestützte Systeme an, die durch maschinelles Lernen in der Lage sind, komplexe Muster zu erkennen, sich an neue Angriffstypen anzupassen und gleichzeitig Fehlalarme zu reduzieren. Auf diese Ansätze werden in den folgenden Kapiteln eingegangen.

2.4 Einführung in Künstliche Intelligenz und Maschinelles Lernen

Mit der rasanten Entwicklung moderner Technologien in den vergangenen Jahren hat sich auch die Künstliche Intelligenz (KI) erheblich weiterentwickelt. Sie ist heute fester Bestandteil des Alltags, etwa in Smart-Home-Geräten, Navigationssystemen oder medizinischen Behandlungen. Auf der Grundlage

der Beobachtungen kann KI anhand ihrer intelligenten Agenten Entscheidungen treffen, Aktionen ausführen und auf Feedback reagieren.

Das ML ist eine zentrale Technologie innerhalb der KI. Hierbei handelt es sich um eine weit verbreitete Methode zur Datenanalyse, bei der Algorithmen in der Lage sind, aus großen Datenmengen Muster zu erkennen, Regeln abzuleiten und Modelle fortlaufend zu optimieren. Das Ziel ist es, auf dieser Basis Vorhersagen zu treffen oder unbekannte Daten zu klassifizieren (Yanqin Cao, 2023).

ML ist ein Teilgebiet der KI und kein eigenständiger Zweig. Es beinhaltet Algorithmen zur Durchführung von Vorhersagen und zur Generierung von Ergebnissen. Das Konzept wurde bereits 1959 von Arthur Samuel geprägt (Naim et al., 2023). ML gilt heute als eine der Schlüsseltechnologien im Bereich des KI-Computings (Yanqin Cao, 2023). Beide Technologien, KI und ML, nutzen Vorhersagetechniken und Algorithmen, zur Entwicklung intelligenter, individualisierter Anwendungen (Naim et al., 2023).

Eine Weiterentwicklung des ML stellt das DL dar. DL basiert auf künstlichen neuronalen Netzwerken und simuliert die Struktur und Funktion menschlicher Neuronen, um Daten zu analysieren und Vorhersagen zu treffen. Es wird als dritte Stufe in der Entwicklung von ML-Technologien betrachtet und ermöglicht die automatische Extraktion komplexer Merkmale und Muster aus sehr großen Datenmengen. Typische Anwendungsfelder sind die Bilderkennung und Spracherkennung (Yanqin Cao, 2023).

In den folgenden Abschnitten wird das ML näher beleuchtet, insbesondere in Bezug auf seine Bedeutung für die Erkennung von DDoS-Angriffen sowie die damit verbundenen Herausforderungen und Grenzen.

2.4.1 Arten von Maschinellem Lernen

Wie in den vorherigen Kapiteln erklärt wurde, ist ML ein Teilgebiet der KI. Diese kann auch in weitere Unterkategorien klassifiziert werden. Diese wären das überwachte Lernen (Supervised Learning), das unüberwachte Lernen (Unsupervised Learning), das semi-überwachte Lernen (Semi-Supervised Learning) sowie das bestärkende Lernen (Reinforcement Learning).

Um den grundlegenden Unterschied zwischen klassischen regelbasierten Systemen und ML zu verdeutlichen, zeigt Abbildung 5 den unterschiedlichen Aufbau beider Ansätze. Während klassische Systeme Eingabedaten mithilfe fest vorgegebener Regeln in Ausgaben überführen, lernen ML-Systeme Modelle direkt aus Eingabe- und Ausgabedaten.

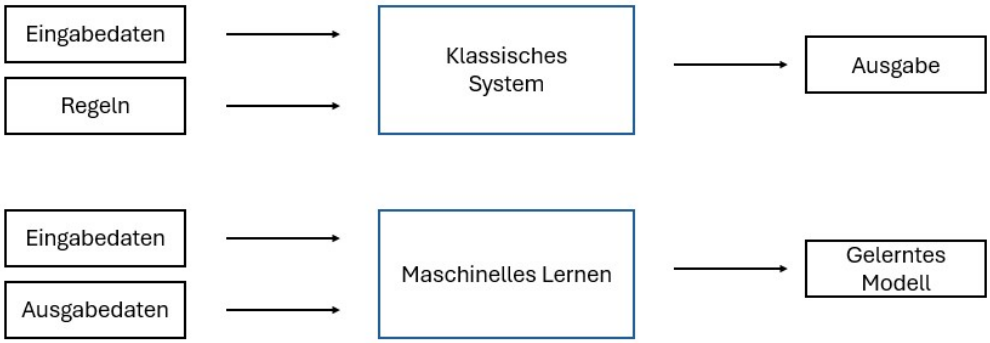

Abbildung 5: Klassisches System vs. Maschinelles Lernen. Eigene Darstellung

Beim überwachten Lernen handelt es sich um eine Form des ML, bei der ein Algorithmus eine Funktion auf Grundlage gelabelter Trainingsdaten erlernt (Rincy und Gupta, 2020). Dieses System wird mit Datensätzen trainiert, bei denen die Zielausgabe bereits bekannt ist. In der Fachliteratur wird dies als labeled data bezeichnet (Gupta et al., 2022).

Das überwachte Lernen hat hierbei das Ziel, aus diesen Trainingsdaten eine Funktion abzuleiten, die in der Lage ist, neue Instanzen korrekt zu klassifizieren (Rincy und Gupta, 2020). Der Lernprozess basiert auf vergangenen Erfahrungen. Das System bedient sich bekannter Eingabe-Ausgabe-Paare, um Prognosen für neue Daten treffen zu können (Gupta et al., 2022).

Es werden zwischen zwei Hauptkategorien unterschieden. Die erste wäre die Regression, welche zur Vorhersage kontinuierlicher Werte dient. Die zweite ist die Klassifikation, bei der Daten in diskrete Klassen eingeteilt werden, wie z.B. „Spam“ oder „kein Spam“. Auch Mehrklassenklassifikationen, wie z.B. „Mann“, „Auto“ oder „Hund“ sind möglich (Gupta et al., 2022).

Gängige Algorithmen des überwachten Lernens sind z.B. Entscheidungsbäume (Decision Trees) und Support Vector Machines (SVM). Sie zeichnen sich durch eine hohe Genauigkeit bei Klassifikationsaufgaben und eine vergleichsweise geringe Fehlerrate aus (Gupta et al., 2022).

Alltägliche Beispielanwendungen wären z.B. in der Bilder- und Spracherkennung, wobei die Maschinen die menschliche Sprache verstehen und danach handeln. Ein weiteres Beispiel wären die Chatbots. Dies sind konversationelle Software-Tools mit menschenähnlichen Abfrageantwortfähigkeiten (Gupta et al., 2022).

Das unüberwachte Lernen unterscheidet sich grundlegend vom überwachten Lernen, da es ohne gelabelte Daten und Anleitungen (Supervision) arbeitet (Gupta et al., 2022; Dalal, 2020). Im Gegensatz zum überwachten Lernen lernt die Maschine selbstständig, ohne vorgegebene Zielwerte oder Ausgaben. Dabei ist das Ziel, verborgene Muster und Strukturen in den Rohdaten zu entdecken (Gupta et al., 2022).

Der Lernprozess erfolgt, indem der Algorithmus Ähnlichkeiten zwischen Datenpunkten identifiziert und diese anschließend in Gruppen (Cluster) einteilt. Ein typisches Beispiel wären unbeschriftete Bilder oder Muster, bei denen keine Kategorie vorgegeben ist. Das System versucht, in diesen Daten sinnvolle Strukturen zu erkennen (Gupta et al., 2022; Dalal, 2020).

Unüberwachtes Lernen hat zwei Hauptkategorien, welche zum einen das Clustering und zum anderen das Assoziationslernen sind. Ein bekannter Clustering-Algorithmus ist bspw. K-Means, bei dem Daten nach Ähnlichkeiten auf Basis geometrischer Distanzen gruppiert werden (Dalal, 2020). Ein klassisches Beispiel für Assoziationslernen ist der Apriori-Algorithmus, der in Transaktionsdatenbanken häufig auftretende Itemsets und deren Zusammenhänge identifiziert (Yuan, 2017).

Anwendungen des unüberwachten Lernens finden sich vor allem in der Analyse großer Datenmengen (Gupta et al., 2022). In IoT-Netzwerken werden unüberwachte Algorithmen eingesetzt, um Lasten auszugleichen, Geräte bzw. Knoten zu clustern und Intrusionen, Fehler und Anomalien zu erkennen (Dalal, 2020).

Auch zur Erkennung von DDoS-Angriffen kann unüberwachtes Lernen beitragen. Da keine gelabelten Daten erforderlich sind, muss das System nicht zwischen „normalem“ und „angreifendem“ Verhalten unterscheiden lernen. Stattdessen erkennt es automatisch auffällige Muster. Ein Beispiel dafür ist die multivariate Korrelationsanalyse, bei der der Algorithmus nach ungewöhnlichen Abweichungen vom erwarteten Netzwerkverhalten sucht. Ein DDoS-Angriff erscheint dabei als deutliches Anomalie-Muster und kann so als potenzielle Bedrohung identifiziert werden (Gupta et al., 2022).

Das semi-überwachte Lernen integriert das überwachte Lernen und unüberwachte Lernen. Dies ist ein Schlüsselaspekt der Netzwerksicherheit, um verschiedene Angriffe im Netzwerk präzise zu identifizieren. Die meisten semi-überwachten Lernverfahren basieren auf Selbstkodierung und gliedern den Lernprozess typischerweise in zwei Phasen. Zunächst in ein unüberwachtes Feature-Lernen, gefolgt von einem überwachten Fein-Tuning (Lu und Ding, 2021). Viele dieser Algorithmen stützen sich auf Annahmen über die Datenstruktur, wie bspw. die Manifold-Annahme, die davon ausgeht, dass die Daten auf einer niedrigdimensionalen Mannigfaltigkeit liegen oder die Cluster Annahme, der zufolge ähnliche Datenpunkte derselben Klasse angehören und Entscheidungsgrenzen durch Regionen geringer Datendichte verlaufen (Mallapragada et al., 2009).

Das Ziel semi-überwachter Lernverfahren besteht darin, die Effizienz und Genauigkeit von Klassifikationsmodellen zu verbessern, indem sowohl eine geringe Menge an gelabelten Daten als auch eine große Menge an ungelabelter Daten genutzt wird. Ein spezifisches Ziel ist das sogenannte semi-supervised improvement, bei dem die Leistungsfähigkeit eines bestehenden überwachten Lernalgorithmus durch den Einsatz ungelabelter Daten erhöht werden soll, ohne dass ein komplett neuer Algorithmus entwickelt werden muss. Außerdem gehört zu den Zielen, ein Modell zu erstellen, das auch mit einer begrenzten Anzahl gelabelter Daten eine hohe Genauigkeit erreicht (Mallapragada et al., 2009).

Semi-überwachtes Lernen kombiniert die Fähigkeit des überwachten Lernens, aus gelabelten Daten zu lernen, mit der Stärke des unüberwachten Lernens, Muster in ungelabelter Daten zu erkennen. Ein wesentlicher Vorteil besteht in der Kosteneffizienz, da der Bedarf an manuell gelabelten, teuren und zeitaufwendigen Daten reduziert wird, indem ungelabelte Daten effektiv genutzt werden. Zudem ermöglicht

die Ausnutzung der zugrunde liegenden Struktur ungelabelter Daten eine präzisere und robustere Bestimmung der Entscheidungsgrenzen von Modellen (Lu und Ding, 2021; Mallapragada et al., 2009).

Beispiele für den Einsatz semi-überwachter Lernverfahren sind etwa Systeme zur Spam-Erkennung, bei denen nur wenige gelabelte E-Mails („Spam“ oder „kein Spam“) benötigt werden, während zusätzlich Hunderte oder Tausende ungelabelter E-Mails zur Erkennung verdächtiger Muster herangezogen werden können (Mallapragada et al., 2009). Ebenso können solche Verfahren, wie von Lu und Ding (2021) beschrieben, zur Erkennung bösartigen Netzwerkverkehrs eingesetzt werden, um auch neue oder sich ständig verändernde Arten von Netzwerkangriffen zu identifizieren, selbst wenn nur eine kleine Menge bekannter bösartiger Verkehrsdaten vorliegt.

Neben den zuvor beschriebenen Lernarten stellt das bestärkende Lernen einen weiteren wichtigen Ansatz innerhalb des ML dar. Es unterscheidet sich von anderen Verfahren dadurch, dass der Agent selbst herausfinden muss, welche Aktionen zur maximalen Belohnung führen, ohne dass ihm vorgegeben wird, was richtig oder falsch ist. Dabei handelt es sich um einen Trial-and-Error-Mechanismus, bei dem der Agent durch ständiges Ausprobieren, Fehler und Rückmeldungen lernt. Dieser Lernprozess ist dem desmenschlichen Lernens sehr ähnlich und konzentriert sich insbesondere auf das Erlernen von Problemlösungsstrategien, was als entscheidend für den Fortschritt in Richtung KI betrachtet wird (Wang, 2020).

Im Kern besteht die grundlegende Interaktion des bestärkenden Lernens darin, dass ein Agent mit einer Umgebung interagiert. Er beobachtet zunächst den Zustand der Umgebung, wählt anschließend eine Aktion, woraufhin die Umgebung einen neuen Zustand sowie ein Belohnungssignal generiert. Ein typisches System dieses Lernverfahrens umfasst dabei vier zentrale Elemente: die Politik, die angibt, welche Aktion in welchem Zustand ausgeführt wird, das Belohnungssignal, die Wertfunktion und ein optionales Umgebungsmodell (Wang, 2020).

Das bestärkende Lernverfahren basiert häufig auf dem Konzept des Markov-Entscheidungsprozesses (MDP), bei dem ein Agent Zustände der Umgebung wahrnimmt, Aktionen auswählt und dafür Belohnungen erhält. Ziel ist es, eine optimale Politik zu entwickeln, die die langfristige Gesamtbelohnung maximiert. Es existieren verschiedene Algorithmen, darunter Q-Learning, das erwartete Belohnungen unabhängig von der aktuellen Politik ermittelt. Sarsa, dass diese auf Basis der tatsächlich genutzten Politik berechnet, sowie Policy-Gradient-Ansätze, die die Politik direkt mithilfe von Gradientenverfahren optimieren (Abou El Houda et al., 2022; Wang, 2020).

Das bestärkende Lernen bietet mehrere Vorteile. Es ermöglicht Lernen durch direkte Interaktion mit der Umgebung, ohne dass vorheriges Wissen über deren Dynamik erforderlich ist. Außerdem hat sich gezeigt, dass dies besonders effizient zur Lösung komplexer Probleme geeignet ist. Darüber hinaus gilt die Kombination von dem bestärkenden Lernen mit anderen Lernmethoden, wie dem überwachten und unüberwachten Lernen, als vielversprechend und fördert die Vereinheitlichung maschineller Lernansätze (Wang, 2020; Abou El Houda et al., 2022).

2.4.2 Relevanz für DDoS-Erkennung

DDoS-Angriffe stellen nach wie vor eine ernsthafte Bedrohung für die Netzwerksicherheit dar. Angesichts der zunehmenden Komplexität und Dynamik dieser Angriffsformen stoßen klassische Sicherheitsmaßnahmen an ihre Grenzen. ML-Techniken haben sich in diesem Kontext als äußerst wirksam erwiesen und werden zunehmend zur Erkennung und Abwehr von DDoS-Angriffen eingesetzt (Naing und Thwel, 2023).

Ein zentraler Vorteil von ML liegt in der Fähigkeit, sowohl bekannte als auch bislang unbekannte Angriffsmuster zu identifizieren. Durch das Training auf Netzwerkverkehrsdaten können ML-Modelle frühzeitig verdächtige Aktivitäten erkennen und so größeren Schaden verhindern (Naing und Thwel, 2023). Da sich DDoS-Angriffe stetig weiterentwickeln und oft keine klaren Vorwarnungen geben, ist der Einsatz lernfähiger, adaptiver Systeme besonders relevant (Raj und Kang, 2022).

Experimente belegen die hohe Genauigkeit und Leistungsfähigkeit von ML-Klassifikatoren bei der Unterscheidung zwischen normalem und bösartigem Datenverkehr. Studien zeigen, dass Verfahren wie Logistic Regression, Random Forest oder XGBoost in vielen Szenarien sehr gute Erkennungsraten erzielen (Naing und Thwel, 2023; Raj und Kang, 2022).

Ein weiterer Vorteil von ML liegt in der Reduktion von Fehlalarmen. Für diesen Zweck kommen häufig semi-überwachte Verfahren zum Einsatz, die wie in Kapitel 2.4.1 beschrieben, die Stärken von überwachtem und unüberwachtem Lernen kombinieren. Ein Beispiel ist das hybride SVM-SOM-Modell, das bekannte Angriffe mithilfe von SVM klassifiziert, während SOM den restlichen Datenverkehr auf unbekannte Muster analysiert. Dieses kombinierte Modell erzielte in Studien eine höhere Genauigkeit und eine niedrige Fehlalarmrate im Vergleich zu einzelnen ML-Modellen (Deepa et al., 2018).

Besonders relevant ist der Einsatz von ML in modernen Netzwerkarchitekturen wie SDN-Netzwerken. Dort übernimmt der zentrale Controller eine kritische Rolle und ist somit ein bevorzugtes Ziel für DDoS-Angriffe. ML-basierte Erkennungsmethoden bieten eine effektive Möglichkeit, den Controller frühzeitig von Angriffen zu schützen (Raj und Kang, 2022; Deepa et al., 2018).

Schließlich ergibt sich der Bedarf an ML-Techniken auch aus den Defiziten klassischer Erkennungsmethoden. Während früher bereits einfache Angriffe, etwa das Überschreiten von ICMP-Spezifikationen oder das gezielte Ausnutzen von TCP/IP-Protokollen ausreichten, um Server lahmzulegen, erfordern heutige Netzwerke angesichts deutlich raffinierterer Angriffsmethoden automatisierte, datenbasierte Verteidigungsstrategien, wie sie ML bietet (Naing und Thwel, 2023; Raj und Kang, 2022).

Ein besonders häufig eingesetzter Ansatz zur Erkennung von DDoS-Angriffen ist das überwachte Lernen. Die Leistung von Klassifikatoren zur DDoS-Erkennung wird dabei anhand verschiedener Bewertungskriterien wie Genauigkeit (Accuracy), Präzision (Precision), Rückruf (Recall) und dem F1-Score bewertet. Diese Metriken basieren auf der sogenannten Konfusionsmatrix, die zwischen folgenden Kategorien unterscheidet (Naing und Thwel, 2023):

- True Positive (TP): Anzahl der bösartigen Anfragen, die korrekt als DDoS-Angriff klassifiziert wurden
- True Negative (TN): Anzahl der legitimen Anfragen, die korrekt als normaler Verkehr erkannt wurden
- False Positive (FP): Anzahl der legitimen Anfragen, die fälschlicherweise als DDoS-Angriffe eingestuft wurden (Fehlalarm)
- False Negative (FN): Anzahl der bösartigen Anfragen, die fälschlicherweise als normaler Verkehr klassifiziert wurden und somit unerkannt blieben (Naing und Thwel, 2023)

Die Genauigkeit beschreibt den prozentualen Anteil der korrekt klassifizierten Datenpunkte am Gesamtdatensatz (Deepa et al., 2018), also sowohl korrekt erkannte Angriffe als auch korrekt erkannten Normalverkehr. Die Präzision zeigt, wie viele der als Angriffe klassifizierten Fälle tatsächlich Angriffe waren (Naing und Thwel, 2023). Der Rückruf (auch Erkennungsrate genannt) gibt an, wie viele der tatsächlichen Angriffe vom Modell korrekt erkannt wurden (Deepa et al., 2018). Der F1-Score stellt das harmonische Mittel von Präzision und Rückruf dar und ist besonders hilfreich bei unausgeglichenen Datenverteilungen, da er sowohl FP als auch FN berücksichtigt (Nain und Thwel, 2023).

Ein häufig genutzter Klassifikationsalgorithmus ist die Logistic Regression, die vor allem für Klassifikationsprobleme eingesetzt wird und auch nichtlineare Zusammenhänge zwischen Variablen abbilden kann (Naing und Thwel, 2023). Naing und Thwel (2023) identifizierten Logistic Regression als den leistungsstärksten Klassifikator für die DDoS-Angriffsklassifizierung und erreichten eine Genauigkeit von 93%. Im Gegensatz dazu ergab eine andere Studie von Raj und Kang (2022), dass Logistic Regression in ihren Experimenten mit anderen Modellen die niedrigste Genauigkeit von 83,84% aufwies. Dies zeigt, dass die Performance stark vom verwendeten Datensatz und der spezifischen Implementierung abhängen kann.

Ein weiterer klassischer Algorithmus ist SVM, der sowohl für Klassifikations- als auch für Regressionsprobleme geeignet ist. Lineare SVMs werden bei linear trennbaren Daten eingesetzt, während nichtlineare Varianten auch komplexere Muster erkennen können. Ein Vorteil von SVM ist die im Vergleich zu neuronalen Netzwerken oft schnellere Verarbeitung und höhere Effizienz. In der Studie von Naing und Thwel (2023) erreichte SVM jedoch lediglich eine Genauigkeit von 77% und wurde daher als „schwacher Lerner“ für den Datensatz eingestuft (Naing und Thwel, 2023).

XGBoost, ein leistungsstarker Gradient-Boosting-Algorithmus, gilt in der angewandten Praxis als besonders effizient. Raj und Kang (2022) erzielten mit XGBoost die höchste Genauigkeit unter allen getesteten Algorithmen (98,38%) und bewerteten ihn daher als besonders geeigneten Klassifikator zur DDoS-Erkennung.

Auch der Random-Forest-Algorithmus, ein ensemblebasiertes Verfahren, wird häufig zur DDoS-Erkennung verwendet. Er kombiniert mehrere Entscheidungsbäume, um die Einschränkungen einzelner Bäume zu umgehen und die Klassifikationsleistung zu verbessern (Garg et al., 2021). Zusätzlich wird

Random Forest auch für die Merkmalsauswahl (Feature Selection) eingesetzt, um die relevantesten Merkmale mit maximalem Einfluss auf das Zielset auszuwählen (Raj und Kang, 2022).

Neben dem überwachten Lernen spielt auch das unüberwachte Lernen eine zentrale Rolle bei der Erkennung von DDoS-Angriffen. Diese Verfahren analysieren den eingehenden Netzwerkverkehr, um Anomalien zu identifizieren, ohne dabei auf zuvor gelabelte Trainingsdaten angewiesen zu sein (Garg et al., 2021).

Ein beliebtes Beispiel für ein solches Verfahren ist die SOM. Das Ziel dieses Algorithmus ist es, hochdimensionale Daten auf eine niedrigdimensionale Karte, meist zweidimensional abzubilden, wobei die ursprünglichen räumlichen Beziehungen zwischen den Datenpunkten erhalten bleiben. Ähnliche Datenpunkte erscheinen somit nahe beieinander, während unterschiedliche weiter voneinander entfernt dargestellt werden. Da SOM keine Annahmen über die Datenverteilung macht, wird es häufig zur Datenvisualisierung oder als Vorbereitungsschritt für weiterführende Analysen wie Clustering eingesetzt (Li und Lin, 2018). In der DDoS-Erkennung sind SOMs besonders nützlich, da sie auch neue, bislang unbekannte Angriffsmuster erkennen können, was sie gegenüber rein signaturbasierten Methoden deutlich flexibler macht (Deepa et al., 2018). Deepa et al. (2018) schlagen zudem ein hybrides Modell namens SVM-SOM vor. Dabei klassifiziert eine SVM zunächst bekannte Angriffe, während der SOM-Anteil anschließend aus dem verbleibenden Netzwerkverkehr neue Angriffe identifiziert. Dieses kombinierte Modell erreichte eine höhere Genauigkeit und Erkennungsrate bei gleichzeitig reduzierter Fehlalarmrate im Vergleich zu den Einzelmodellen. Im Gegensatz dazu kann bei alleiniger Nutzung von SOM die False Positive Rate (FPR) relativ hoch sein.

Ein weiteres Verfahren ist das Co-Clustering. Dies ist ein Algorithmus, der gleichzeitig Zeilen und Spalten von Daten gruppiert. Im Kontext der Netzwerkverkehrsanalyse kann dies bedeuten, dass nicht nur ähnliche Datenströme, sondern auch ähnliche Merkmale der Datenströme gruppiert werden. Im Rahmen der DDoS-Erkennung wird Co-Clustering häufig als unüberwachter Bestandteil semi-überwachter Ansätze verwendet. Ziel ist es, den Datenumfang zu reduzieren, indem normaler und irrelevanter Verkehr ausgeschlossen wird. Dies ist besonders wichtig, da unbekannter Normalverkehr oft zu einer erhöhten FPR führt. Ein konkreter Ansatz nutzt dabei einen spektralen Co-Clustering-Algorithmus, der den Netzwerkverkehr in drei Hauptcluster aufteilt. Durch Ausschluss des als „normal“ identifizierten Clusters kann der Klassifikationsaufwand reduziert und die Genauigkeit verbessert werden (Garg et al., 2021).

Der MeanShift-Algorithmus stellt einen unüberwachten, nicht parametischen Clustering-Ansatz dar, der lokale Dichtemaxima in den Daten identifiziert. In Bezug auf DDoS-Angriffe kann MeanShift zur Erkennung von Angriffen in Offline-Datensätzen eingesetzt werden (Raj und Kang, 2022).

Ein weiterer weit verbreiteter Algorithmus ist K-Means, der eine Datenmenge in eine vordefinierte Anzahl von Clustern aufteilt. Dabei werden Datenpunkten den jeweils nächstgelegenen Mittelpunkt (Zentroiden) zugewiesen. K-Means ist in der DDoS-Erkennung nützlich, da es hilft, verkehrsbasierte

Muster zu gruppieren und damit potenziell schädliche Datenströme zu identifizieren (Deepa et al., 2018).

Ein verwandter Algorithmus ist K-Medoids, der im Gegensatz zu K-Means einen tatsächlichen Datenpunkt als Zentrum (Medoid) verwendet, was ihn robuster gegenüber Ausreißern macht. Auch K-Medoids wurden zur Erkennung von DDoS-Angriffen eingesetzt, insbesondere in Kombination mit anderen Clustering- oder Klassifikationsverfahren (Deepa et al., 2018; James et al., 2024).

Insgesamt bieten unüberwachte Lernmethoden wie SOM, Co-Clustering, MeanShift, K-Means und K-Medoids leistungsstarke Werkzeuge zur Erkennung unbekannter Angriffsmuster. Sie tragen zur Verbesserung der Datenverarbeitungseffizienz bei und werden häufig in Kombination mit überwachten Verfahren eingesetzt, um die Detektionsgenauigkeit zu erhöhen und Fehlalarme zu reduzieren (Deepa et al., 2018).

Auch das bestärkende Lernen wird zunehmend in der DDoS-Erkennung eingesetzt, insbesondere im Bereich der IoT-Netzwerke, die durch ihre begrenzten Ressourcen besonders anfällig für Angriffe sind. Moderne Ansätze wie Multi-Agent-Deep-Reinforcement-Learning (MADRL) in Kombination mit einem SDN-Controller zeigen, dass sich durch adaptive Lernprozesse Angriffsmuster effektiv erkennen und Abwehrmaßnahmen automatisch einleiten lassen. (Shaharkar et al., 2024; James et al., 2024).

Diese Vielzahl an Verfahren und Einsatzmöglichkeiten verdeutlicht die zentrale Rolle, die ML bei der effektiven Erkennung von DDoS-Angriffen spielt.

2.4.3 Herausforderungen und Grenzen

DDoS-Angriffe stellen eine dynamische Herausforderung im Bereich der Netzwerksicherheit dar, da sie in Größe, Volumen und Art der verwendeten Ressourcen stark variieren können. ML bietet grundsätzlich vielversprechende Ansätze zur Erkennung solcher Angriffe, stößt jedoch in der Praxis auf zahlreiche Einschränkungen. Ein zentrales Problem liegt darin, dass ML-Methoden auf zuvor gelernten Angriffsmustern basieren, wodurch sie Schwierigkeiten haben, neuartige oder abgewandelte Angriffsformen zuverlässig zu erkennen (Jyoti und Behal, 2021).

Die Erstellung gelabelter Datensätze erfordert zudem erhebliche zeitliche Ressourcen. Gleichzeitig kann das Vorhandensein großer Mengen irrelevanter Daten die Leistungsfähigkeit von IDS erheblich beeinträchtigen (Jyoti und Behal, 2021). Insbesondere im Kontext von IoT-Systemen ergeben sich zusätzliche Herausforderungen aufgrund beschränkter Rechenkapazitäten sowie des Bedarfs an umfangreichen Trainingsdaten. Bei einem DDoS-Angriff kann es außerdem zu erheblichem Datenverlust kommen, welcher sich negativ auf die Klassifizierung der Modelle auswirkt (Wehbi et al., 2019).

Ein weiteres Problem ist die hohe FPR vieler ML-Methoden. Viele Ansätze konzentrieren sich auf die Erkennung von nicht-legitimem Datenverkehr. Dabei wird jedoch häufig legitimer Datenverkehr fälschlicherweise klassifiziert. Dies führt zu eine unbeabsichtigten Leistungsverschlechterung des regulären

Netzwerkverkehrs (Wehbi et al., 2019). Überwachte Algorithmen sind besonders stark auf gelabelte Netzwerkdaten angewiesen (Garg et al., 2021) und können bei großen, rauschbehafteten Datenmengen an ihre Grenzen stoßen. Unüberwachte Verfahren wiederum weisen typischerweise eine hohe FPR auf, was legitime Nutzer blockieren, und die Netzwerknutzung einschränken kann (Jyoti und Behal, 2021; Garg et al., 2021). Auch unbekannte Muster im normalen Verkehr tragen zur Erhöhung des FPR bei und beeinträchtigen die Klassifikationsgenauigkeit zusätzlich (Garg et al., 2021).

Ebenso auf der Ebene der Algorithmen werden Schwächen deutlich. Random Forest erzielt zwar in vielen Studien eine sehr hohe Erkennungsrate, erweist sich jedoch aufgrund der vergleichsweisen langen Trainingszeit als ineffizient für Echtzeitanwendungen (Jyoti und Behal, 2021; Webi et al., 2019). SVMs gelten als schwer interpretierbar und zeigen in manchen Szenarien eine geringere Genauigkeit, während K-Nearest-Neighbor (KNN) durch sein zeitaufwändiges Rechenverhalten problematisch ist, insbesondere bei großen Datenmengen (Wehbi et al., 2019).

Beim Einsatz vom bestärkenden Lernen kommen weitere Herausforderungen hinzu. Die Generalisierbarkeit dieser Modelle ist häufig begrenzt. Viele dieser Ansätze können sich nur schwer an unbekannte oder neue Szenarien anpassen. Zudem ist ein weiteres bestehendes Problem das Datenungleichgewicht (Class Imbalance). Dies können den Lernprozess und die Fähigkeit der Systeme, seltene, aber kritische Anomalien zu erkennen, negativ beeinflussen. Hinzu kommt, dass viele Modelle des bestärkenden Lernens auf manuell definierten Belohnungsfunktionen basieren, die zu Verzerrungen führen und die Lerneffizienz einschränken können. Auch Probleme wie die Überbewertung des Q-Wertes können zu fehlerhaften Klassifikationen führen. Ebenso kann die unzureichende Exploration eine optimale Richtlinienfindung verhindern (Al-Fawa'reh et al., 2024).

Semi-überwachtes Lernen stellt zwar einen vielversprechenden Ansatz dar, da es die Vorteile von überwachtem und unüberwachtem Lernen kombiniert, bringt jedoch eigene Herausforderungen mit sich. Semi-überwachtes IDS erfordert eine hochentwickelte Implementierung, um die Integration gelabelter und ungelabelter Daten effektiv zu ermöglichen (Jyoti und Behal, 2021; Garg et al., 2021).

Ein besonders schwerwiegender Spezialfall stellen Zero-Day-Angriffe dar. Dabei handelt es sich um neuartige und bisher unbekannte DDoS-Angriffe, für die keine passenden Trainingsdaten vorliegen (Minhas et al., 2025). Sie stützen sich nicht auf bekannte Signaturen oder Verhaltensmuster und können daher in jede Kategorie von DDoS-Angriffen fallen. Dies erschwert ihre Vorhersehbarkeit und stellt die Generalisierungsfähigkeit bestehender ML-Modelle infrage (Al-Fawa'reh et al., 2024). Zero-Day-Angriffe können zudem so unauffällig gestaltet sein, dass sie unterhalb von Anomalie-Erkennungsschwellen bleiben. Auch signaturbasierte IDS stoßen hier an ihre Grenzen, da es unmöglich ist, alle potenziellen neuen Angriffssignaturen im Voraus zu erfassen (Minhas et al., 2025).

Insgesamt zeigt sich, dass KI-gestützte Methoden sowohl vielversprechend als auch herausfordernd sind, insbesondere im Umgang mit dynamischen Angriffen wie Zero-Day-Angriffe. Um das Potenzial

dieser Technologien besser einordnen zu können, folgt im nächsten Abschnitt eine Betrachtung der Rolle von KI in der übergeordneten IT-Sicherheitsstruktur.

2.5 Rolle von KI in der IT-Sicherheitsarchitektur

Um die Rolle von KI innerhalb der IT-Sicherheitsarchitektur besser zu verstehen, ist zunächst ein Überblick über ihre allgemeinen Einsatzbereiche in sicherheitsrelevanten Systemen erforderlich.

2.5.1 Allgemeine Bedeutung von KI in der IT-Sicherheit

KI bezeichnet den Versuch, menschenähnliche Entscheidungsstrukturen in einem nicht eindeutigen Umfeld nachzubilden. Ziel ist es, Computersysteme zu befähigen, eigenständig Probleme zu bearbeiten und Entscheidungen zu treffen. Dabei lernt das System aus Beispielen, sogenannten Lerndaten, erkennt Muster sowie Gesetzmäßigkeiten und wendet diese anschließend auf neue Konstellationen, also auf Nutzdaten an. Im Bereich des ML wird dieser Mechanismus besonders deutlich (siehe Abbildung 5). Beim Anlernen erhält das System das Ergebnis, die zugrunde liegenden Kriterien muss es jedoch selbst erschließen (von Faber und Kohler, 2019).

Grundsätzlich wird zwischen schwacher und starker KI unterschieden. Die schwache KI wird mit ML gleichgesetzt und auf spezifische Aufgaben beschränkt, während die sogenannte starke KI (Artificial General Intelligence, AGI) das Ziel verfolgt, menschenähnliche Intelligenz vollständig zu reproduzieren. Diese hypothetische Form der KI könnte die menschliche Intelligenz sogar übertreffen und sich eigenständig weiterentwickeln, was potenziell zu unvorhersehbaren Fortschritten und einem Kontrollverlust führen könnte (Pohlmann, 2025).

Zu den jüngsten Entwicklungen zählt das Large Language Model (LLM), das als Grundlage für generative KI (GenAI) dient. Diese Systeme sind in der Lage, basierend auf großen Mengen an Trainingsdaten eigenständige Texte, Bilder oder Code zu generieren, wie es etwa bei ChatGPT der Fall ist. Werden sie entsprechend trainiert, können LLMs sogar den individuellen Schreibstil einer bestimmten Person imitieren (Pohlmann, 2025). Während AGI und LLMs im Kontext dieser Arbeit nicht im Vordergrund stehen, verdeutlicht ihre Erwähnung die Breite aktueller KI-Entwicklungen, die im Ausblick (Kapitel 6) noch einmal aufgegriffen werden.

Der Einsatz von KI in der IT-Sicherheitsstruktur bietet erhebliche Vorteile und Mehrwerte für Unternehmen und Organisationen. Ein zentrales Anwendungsfeld ist die Erhöhung der Erkennungsrate von Angriffen. KI-gestützte Systeme können sicherheitsrelevante Aktivitäten in Netzwerken, an Endgeräten, auf Servern sowie bei IoT-Geräten und Cloud-Anwendungen analysieren und potenzielle Bedrohungen erkennen. Dazu zählen unter anderem die Erkennung von Schadsoftware (Anti-Malware), Spam, Fake News oder Deepfakes (von Faber und Kohler, 2019; Pohlmann, 2025).

Darüber hinaus trägt KI zur Entlastung von IT-Sicherheitsexperten bei. Sie übernimmt zeitaufwendige Analysearbeiten und kann Prozesse automatisieren. So sind KI-Systeme etwa in der Lage, eine Vielzahl sicherheitsrelevanter Ereignisse gleichzeitig zu analysieren und deren Priorität festzulegen. Dies ermöglicht eine effizientere Bearbeitung durch menschliche Experten. In bestimmten Fällen können KI-Systeme auch (teil-)autonome Reaktionen auslösen, bspw. durch automatische Anpassung von Firewall- oder E-Mail-Regeln im Falle eines Angriffs. Auf diese Weise kann der Schutz zentraler Unternehmensprozesse aufrechterhalten und die Angriffsfläche reduziert werden (Pohlmann, 2025).

Die allgemeine Einbettung von KI in die IT-Sicherheitsarchitektur zeigt bereits ihr großes Potenzial zur Unterstützung sicherheitsrelevanter Prozesse. Doch welche konkreten Vorteile ergeben sich aus dem Einsatz von KI in der IT-Sicherheit? Der folgende Abschnitt beleuchtet die wesentlichen Stärken von KI-basierten Sicherheitslösungen.

2.5.2 Vorteile KI-gestützter Sicherheitslösungen

KI bietet zahlreiche Vorteile für die IT-Sicherheitsarchitektur, insbesondere bei der Erkennung neuer und komplexer Angriffsarten. Durch die Analyse riesiger Datenmengen ist KI in der Lage, Muster und Zusammenhänge zu identifizieren, die für den Menschen kaum mehr nachvollziehbar wären. Dies ist entscheidend, um subtile oder neue Angriffsmuster zu erkennen, die sich in großen Datenströmen verbergen (Kitzmann, 2022).

Des Weiteren ist ein wesentlicher Vorteil der KI, schnellere und präzisere Problemanalysen durchzuführen und zuverlässige Vorhersagen zu treffen. Dies ist für eine effektive Sicherheitslösung von großer Bedeutung (Merkel-Kiss und von Garrel, 2022). Besonders im Bereich der Cyber-Security finden KI-Anwendungen zunehmend Einsatz. Automatisierte Alarm- und Überwachungssysteme, die auf KI basieren, ermöglichen eine rasche Aufdeckung von Missbrauchsfällen (Kitzmann, 2022).

Darüber hinaus verbessert KI das Risikomanagement, etwa im Bereich der Prävention und Verteilungskontrolle. Sie trägt zur Automatisierung sicherheitsrelevanter Prozesse bei, reduziert menschliche Fehler und steigert somit die Qualität und Effizienz der Abläufe, was auch zu einer Kostenreduktion führen kann (Merkel-Kiss und von Garrel, 2022; Kitzmann, 2022).

KI-Systeme haben gegenüber klassischen Sicherheitsmodellen einen entscheidenden Vorteil, dies wäre ihre selbstlernende Fähigkeit. Sie können eigenständig Lernprozesse einleiten, sich ohne menschliches Eingreifen weiterentwickeln und aus vergangenen Strukturen ableiten. Insbesondere DL-Prozesse ermöglichen eine zunehmend präzise Analyse großer Datenressourcen und machen dabei bislang unsichtbare Muster und Zusammenhänge sichtbar (Kitzmann, 2022).

Trotz dieser zahlreichen Vorteile und Einsatzmöglichkeiten von KI in der IT-Sicherheitsarchitektur sind mit ihrem Einsatz auch ernstzunehmende Herausforderungen verbunden, welche im folgenden Abschnitt näher betrachtet werden.

2.5.3 Herausforderungen bei der Integration von KI in Sicherheitsarchitekturen

Der Einsatz von KI in der IT-Sicherheit bringt neben zahlreichen Vorteilen auch eine Reihe neuer Herausforderungen mit sich. KI-basierte Systeme verändern die Grundlage klassischer Sicherheitsverfahren, da etablierte Ansätze häufig nur eingeschränkt funktionieren. Die heutigen Konzepte der IT-Sicherheit beruhen meist auf bekannten und relativ stabilen Informationsflüssen. KI hingegen steuert diese Prozesse selbstständig und teilweise unvorhersehbar. Somit stehen Angreifer den Verteidigern nicht mehr zwangsläufig direkt gegenüber. Stattdessen können sie die IT-Sicherheit indirekt beeinträchtigen, bspw. durch die Manipulation von Nutzdaten, die ein KI-basiertes Sicherheitssystem zum Anlernen verwendet (von Faber und Kohler, 2019).

Ein zentrales Problem stellt die Datenqualität und -integrität dar. KI-Systeme funktionieren nur zuverlässig, wenn Trainings- und Nutzdaten qualitativ übereinstimmen und die zu erkennenden Marken enthalten. Eine schlechte Datenqualität führt entsprechend zu schlechten Ergebnissen. Daher muss die Integrität sowohl der Lerndaten als auch des gesamten Lernvorgangs gewährleistet sein, um Manipulationen durch Angreifer zu verhindern (von Faber und Kohler, 2019; Pohlmann, 2025).

Zudem stoßen klassische Sicherheitssysteme an ihre Grenzen. Aktive Sicherheitselemente wie Firewalls und Identitäts- und Zugriffsmanagement können bei KI-gesteuerten IT-Systemen zu Engpässen führen, da Administratoren den sich veränderten Bedarf häufig nicht schnell genug antizipieren können. Auch klassische IDS/IPS-Systeme sind nur begrenzt geeignet, da sich Anomalien durch das sich wandelnde Nutzungsverhalten KI-gesteuerter Systeme permanent verändern. Eine Anpassung dieser Systeme ist schwierig, wenn KI-Anwendungen und Datenströme automatisch gesteuert werden (von Faber und Kohler, 2019).

Ein weiteres Risiko liegt im Missbrauch von KI durch Angreifer. Diese nutzen KI, um ihre Angriffe effizienter und gezielter zu gestalten. So kann etwa KI verwendet werden, um Schwellenwerte oder typische Muster von Angriffserkennungssystemen zu analysieren und dadurch Angriffe unentdeckt durchzuführen. Im Bereich des Social Engineerings kommen automatisierte Methoden wie Social Bots zum Einsatz, die Fake Accounts und -Nachrichten generieren, um Menschen gezielt zu manipulieren. KI ist zudem in der Lage, künstliche Fotos und Profile zu erstellen. Große Sprachmodelle wie LLMs vereinfachen die Spear-Phishing- und CEO-Fraud-Angriffe für Angreifer, insbesondere in Kombination mit Audio-Imitationen und Deepfake-Videos (Pohlmann, 2025).

Gleichzeitig wird auch die KI selbst zum Angriffsziel. Angreifer versuchen, Trainingsdaten, Inputdaten, Algorithmen und Modelle zu manipulieren, um falsche Ergebnisse zu provozieren. Bei sogenannten Poisoning-Angriffen fügen sie gezielt bösartige Beispiele in die Trainingsdaten ein, um etwa legitime E-Mails als Spam zu klassifizieren oder umgekehrt. Bei Evasion-Angriffen wiederum gestalten sie Eingaben („Adversarial Examples“) so, dass sie für Menschen unauffällig wirken, aber von KI-Systemen falsch klassifiziert werden, um Sicherheitsmechanismen zu umgehen (Pohlmann, 2025).

Ein grundlegendes Problem bleibt die Black-Box-Natur von KI-Systemen. Deren innere Funktionsweise ist häufig nicht vollständig durchschaubar oder einfach erklärbar, was tiefgreifende Auswirkungen auf die IT-Sicherheit hat und etablierte Sicherheitskonzepte infrage stellt. Insbesondere ML-Modelle lernen aus Beispielen, indem sie Muster und Gesetzmäßigkeiten selbstständig extrahieren, ohne dass ihnen explizite Kriterien vorgegeben werden. Die Anzahl der beim Training bestimmten Parameter kann extrem hoch sein, mitunter mehr als 100 Millionen. Dadurch wird das Gesamtsystem kaum mehr überblickbar. Es bleibt unklar, nach welchen Kriterien oder Mustern das System seine Entscheidungen trifft und welche Eingabeparameter dabei tatsächlich relevant waren, sichtbar ist nur das Ergebnis (Von Faber und Kohler, 2019).

Die genannten Schwächen und Risiken verdeutlichen, dass der Einsatz von KI in der IT-Sicherheit nicht nur neue Herausforderungen mit sich bringt, sondern auch gezielte Schutzmechanismen erfordert, insbesondere im Hinblick auf komplexe und schwer erkennbare Bedrohungen wie DDoS-Angriffe. Vor diesem Hintergrund gewinnt die zuverlässige Erkennung von DDoS-Attacken eine zentrale Bedeutung innerhalb der gesamten IT-Sicherheitsarchitektur, die im folgenden Abschnitt näher betrachtet wird.

2.5.4 Bedeutung für die DDoS-Erkennung im Kontext der Gesamtarchitektur

Die Erkennung von DDoS-Angriffen nimmt im Rahmen der gesamten IT-Sicherheitsarchitektur eine zentrale Rolle ein. Insbesondere der Schutz der persönlichen Daten sowie die Absicherung cloudbasierter Systeme sind von wachsender Relevanz aufgrund der zunehmenden Vielfalt und Häufigkeit von Cyberangriffen. DDoS-Angriffe können schwerwiegende Auswirkungen auf die Verfügbarkeit von Diensten haben und das Vertrauen der Nutzer in digitale Infrastrukturen erheblich beeinträchtigen (Akgun et al., 2022).

Auch die Sicherung der zentralisierten Strukturen gewinnt zunehmend an Bedeutung, da viele Unternehmen aus Kostengründen auf zentralen Cloud-Diensten wie Google Cloud, Microsoft Azure oder Amazon zurückgreifen (Akgun et al., 2022). Innerhalb solcher Architekturen ist eine schnelle und präzise Identifikation bösartiger Datenströme erforderlich, um zeitnah Gegenmaßnahmen einleiten zu können. Moderne Ansätze der DDoS-Erkennung, etwa auf Basis der DL-Modelle, ermöglichen diese Echtzeiterkennung und tragen somit wesentlich zur Minimierung der potenziellen Schäden bei (Hnamte et al., 2024).

Besonders relevant wird die DDoS-Erkennung in SDN-Architekturen. Wie bereits in Kapitel 2.2 erläutert, trennen SDN-Architekturen Steuerungs- und Datenebene, wodurch insbesondere der Controller zu einem bevorzugten Ziel für Angriffe wird. Durch unbefugte Zugriffe oder DoS-Angriffe kann das gesamte Netzwerk kompromittiert werden. Gleichzeitig erschweren bestehende Herausforderungen wie die Abhängigkeit von der Netzwerktopologie, eine unvollständige Angriffserkennung, veraltete

Datensätze sowie hohe Hardwareanforderungen eine effektive Verteidigung gegen neue Bedrohungen (Hammadeh et al., 2024).

Auch im Bereich des IoT spielt die DDoS-Erkennung eine zentrale Rolle, da zahlreiche vernetzte Geräte leicht kompromittiert und für Angriffe missbraucht werden können (Roopak et al., 2020). Die Details hierzu werden im IoT-Unterkapitel (2.2.1) näher betrachtet.

ML-basierte Verfahren, insbesondere DL-Ansätze, bieten vielversprechende Möglichkeiten zur Erkennung von Anomalien und Angriffsmustern in komplexen Netzwerkumgebungen. Sie analysieren den Netzwerkverkehr und identifizieren komplexe Muster. Dies bietet den Vorteil einer sehr hohen Erkennungsgenauigkeit und die Fähigkeit, neue Angriffsarten zu erkennen, was zu einer überlegenen Gesamtleistung gegenüber traditionellen Methoden führt (Roopak et al., 2020; Hnamte et al., 2024). Damit wird der grundlegende Unterschied deutlich. Während klassische Sicherheitssysteme auf fest definierten Regeln basieren, sind KI-gestützte Verfahren in der Lage, Muster selbstständig zu lernen und sich flexibel an neue Angriffsszenarien anzupassen.

Ein Beispiel für den erfolgreichen Einsatz von KI in der DDoS-Erkennung ist das von Akgun et al. (2022) entwickelte Convolutional Neural Network (CNN)-basierte IDS, das mit dem CIC-DDoS2019-Datensatz trainiert wurde und sehr hohe Erkennungsraten bei niedrigen Fehlerraten erzielte.

Auf Basis der zuvor dargestellten Grundlagen wird erkennbar, dass KI-gestützte Methoden besonders vielversprechend sind, da sie durch ihre Lern- und Anpassungsfähigkeit bestehende Defizite kompensieren und die DDoS-Erkennung innerhalb der Gesamtarchitektur erheblich stärken können. Im folgenden Kapitel wird daher untersucht, wie konkrete KI-Modelle zur Erkennung und Abwehr von DDoS-Angriffen eingesetzt werden können.

3 DDoS-Erkennung und -Abwehr durch KI

Klassische Sicherheitssysteme stehen vor großen Herausforderungen durch die zunehmende Komplexität und Dynamik moderner DDoS-Angriffe. Um diese Bedrohungen wirkungsvoll abzuwehren, kommen zunehmend KI-basierte Ansätze zur Erkennung und Abwehr zum Einsatz. ML- und DL-Ansätze bieten eine effektive Bekämpfung von DDoS-Angriffen, da sie die Fähigkeit haben, Muster im Netzwerkverkehr zu analysieren und Anomalien zu erkennen. In diesem Kapitel werden die Anforderungen an solche Erkennungssysteme sowie konkrete KI-Methoden zur Angriffserkennung und -abwehr vorgestellt und bewertet.

3.1 Anforderungen an Erkennungssystemen

Ein effektives DDoS-Erkennungssystem muss eine Vielzahl technischer und wirtschaftlicher Anforderungen erfüllen, um sowohl die Angriffsabwehr als auch die Dienstverfügbarkeit sicherzustellen. Dabei

sind insbesondere die Minimierung von Kosten, die Kontrolle von Ressourcen sowie die schnelle und zielgerichtete Reaktion auf Angriffe von zentraler Bedeutung (Somani et al., 2017).

Ein wichtiger Aspekt ist die Kosten- und Nachhaltigkeitsfrage. Die Kosten für die Abwehr von DDoS-Angriffen sollten stets unterhalb des zur Verfügung stehenden Budgets liegen. Zudem müssen sie geringer sein als die potenziellen Verluste, die während eines Angriffs entstehen könnten (Somani et al., 2017).

Auch Autoscaling und Ressourcenmanagement spielen eine zentrale Rolle. Während das dynamische Hinzufügen oder Entfernen von Ressourcen grundsätzlich hilfreich ist, kann es im Kontext von DDoS-Angriffen zu einem Nachteil werden. So können legitime Autoscaling-Mechanismen durch Angriffe missbraucht werden, was zu einem betrügerischen Ressourcenverbrauch und hohen finanziellen Verlusten führen kann. Daher sollte ein Erkennungssystem in der Lage sein, den zusätzlichen Ressourcenbedarf zu minimieren und eine ressourcenkonfliktfreie Ausführung der Abwehrmaßnahmen selbst bei laufenden Angriffen zu gewährleisten (Somani et al., 2017).

Ein weiterer wesentlicher Punkt ist die Angriffsfilterung. Ziel ist es, die Anzahl der Angriffsanfragen, die den Dienst erreichen, zu minimieren. Dabei reicht ein reiner Traffic-Filtermechanismus nicht aus, da moderne und ausgeklügelte Angriffsformen ihre Merkmale variieren können, um unentdeckt zu bleiben. Hinsichtlich der Verfügbarkeit und Dienstqualität muss das System sicherstellen, dass der Dienst auch während eines Angriffs verfügbar bleibt. Die Dienstqualität sollte auf einem akzeptablen Niveau gehalten werden, sodass legitime Benutzer mit minimaler oder gar keiner Ausfallzeit weiterhin auf den Dienst zugreifen können (Somani et al., 2017).

Ein weiterer kritischer Faktor ist die Minimierung der Abwehrdauer, also der Mitigation Throughput Time (MTT). Diese beschreibt die Zeitspanne zwischen Beginn des Angriffs und dessen vollständiger Abwehr. Eine möglichst kurze MTT hilft, die Gesamtkosten und die Auswirkungen des Angriffs zu reduzieren (Somani et al., 2017).

Schließlich muss das Erkennungssystem auch Maßnahmen zur Schadensbegrenzung beinhalten. Dabei geht es insbesondere darum, Kollateralschäden auf nicht betroffene Dienste oder zu minimieren. Dies kann unter anderem durch die Isolation und Überwachung von Komponenten wie Hypervisoren und Netzwerken erreicht werden (Somani et al., 2017).

3.2 Erkennung durch KI

DDoS-Angriffe in IoT-Netzwerken zählen zu den gefährlichsten Bedrohungen der heutigen Cybersicherheitslandschaft (Bala und Behal, 2024). Aufgrund der großen Angriffsfläche, die durch zahlreiche vernetzte Geräte entsteht, sind klassische Schutzmechanismen oft unzureichend. In diesem Kontext bie-

tet der Einsatz von KI ein vielversprechendes Potenzial. KI kann große Datenmengen effizient analysieren, Muster erkennen und sich an neue, bisher unbekannte Bedrohungen anpassen. Insbesondere bei der frühzeitigen Erkennung von DDoS-Angriffen zeigt sie hohe Effektivität (Salem et al., 2024).

Ein typisches KI-gestütztes Erkennungssystem für DDoS-Angriffe besteht aus mehreren Phasen. Der erste Schritt ist die Datenerfassung und -vorverarbeitung, bei der Daten bereinigt, skaliert und in eine für das Training geeignete Form gebracht werden. Anschließend erfolgt die Merkmalsextraktion und -auswahl, die entscheidend dazu beiträgt, die Trainingszeit zu verkürzen und die Modellgenauigkeit zu verbessern. Hierbei können unterschiedliche Verfahren eingesetzt werden, darunter filterbasierte Methoden (z.B. Korrelationen), wrapperbasierte Ansätze (z.B. Mutual Information) oder embedded Methoden wie Random Forests, die Merkmalsauswahl in das Modelltraining integrieren (Saha et al., 2023).

Im nächsten Schritt werden auf den bereinigten und ausgewählten Merkmalen KI-Modelle aus dem Bereich des ML trainiert. Die Leistungsbewertung solcher Systeme erfolgt üblicherweise anhand gängiger Metriken wie Genauigkeit, Präzision, Rückruf, F1-Score, FPR sowie der Gesamtausführungszeit (Saha et al., 2023).

Zur Umsetzung solcher Systeme werden gängige Programmiersprachen wie Python genutzt sowie Frameworks wie TensorFlow, die speziell für das Training und Testen von ML- und DL-Modellen entwickelt wurden (Bala und Behal, 2024).

Um die praktische Anwendung der beschriebenen Methoden zu veranschaulichen, werden im Folgenden konkrete KI-Modelle zur Erkennung von DDoS-Angriffen vorgestellt.

3.3 Beispiel KI-Modelle

Für die DDoS-Erkennung lassen sich die in der Literatur eingesetzten KI-Modelle grob in zwei Gruppen einteilen. Klassische Verfahren des ML und moderne DL-Ansätze. Im Folgenden werden exemplarische Modelle beider Kategorien vorgestellt und miteinander verglichen.

Ein klassisches ML-Modell ist die SVM, welche auf einem überwachten Lernalgorithmus basiert. Für die DDoS-Erkennung wurde die SVM mit verschiedenen Datensätzen trainiert. RT et al. (2014) nutzten den DARPA-1998- und 2000-Datensatz, wobei der Datenverkehr in Trainings- und Testdaten aufgeteilt wurde. Das Modell konnte Angriffe zuverlässig erkennen, zeigte eine hohe Präzision auch bei wenigen Trainingsbeispielen, reduzierte FPR und klassifizierte unbekannte Datenproben korrekt. Die Ergebnisse lagen bei 95,11% Genauigkeit und einer FP-Rate von 0,008.

Ein weiterentwickelter Ansatz wurde von Li et al. (2018) vorgestellt, die ein zweistufiges Verfahren zur DDoS-Erkennung in SDN-Netzwerken nutzten. In der ersten Phase wurden wichtige Merkmale wie Quell- und Ziel-IP, Quell- und Ziel-Port sowie Protokolltyp extrahiert und deren Entropiewerte berechnet. Auffällige Abnahmen deuteten auf Angriffe hin. Diese Werte wurden anschließend in einem One-Class-SVM-Modell trainiert unter der Verwendung des DARPA-1999-Datensatz. In der zweiten Phase

wurden neue Testpakete oder Switch-Ports in Echtzeit überprüft. Dadurch konnten Angriffe effizient erkannt und in vielen Fällen bereits während des Angriffs eingedämmt werden.

Im Vergleich dazu konzentrierten sich RT et al. (2014) auf die klassische Trennung von normalem und bösartigem Datenverkehr. Die Ergebnisse zeigten, dass beide Ansätze hohe Erkennungsleistungen erreichten. RT et al. erzielten 95,11% Genauigkeit, während Li et al. (2018) mit ihrem zweistufigen Verfahren einen F1-Score von 97% erreichten. Besonders hervorzuheben ist die extrem niedrige FPR von RT et al. sowie die Echtzeitfähigkeit des Modells von Li et al. Im Vergleich zu anderen ML-Modellen wie KNN oder Random Forest erwies sich SVM hier als besonders zuverlässig und effizient.

Auch in Abbildung 6 von Li et al. wird ersichtlich, dass SVM im Gegensatz zu den anderen Modellen am besten abgeschnitten hatte. Mit einer Erkennungsrate von 97% bei 6773 Erkennungen hat es somit die beste Performance. Auch wenn KNN eine höhere Erkennungsrate von 98% aufweist, hat SVM besser abgeschnitten, denn die Präzisionsrate von SVM beträgt 100%. Dies bedeutet, dass alles, was SVM als positiv klassifiziert hat, auch wirklich positiv ist und keine FPR vorliegen. Hingegen beträgt die Präzisionsrate von KNN nur 62%, daher hat es auch viele FPR und somit ist auch die Genauigkeit beeinträchtigt.

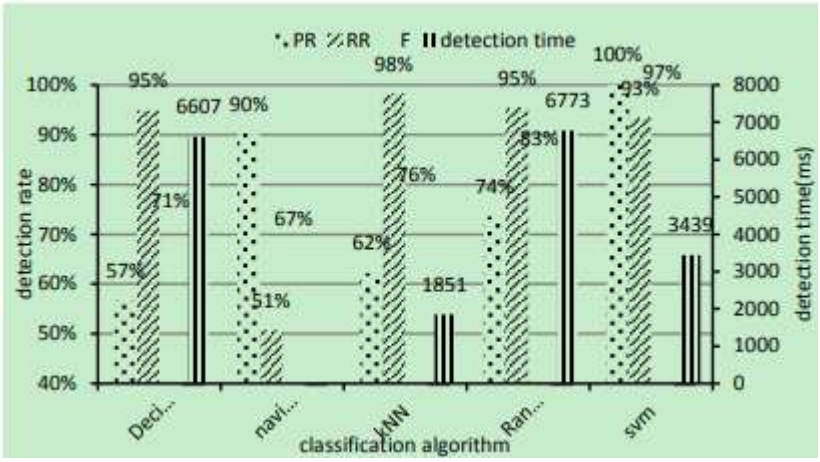

Abbildung 6: Comparison of different algorithms (Quelle: Li et al., 2018, S.7)

Ein weiteres leistungsfähiges Klassifikationsverfahren ist der Random Forest. Dieses Ensemble-Verfahren kombiniert die Vorhersagen mehrerer Entscheidungsbäume, um die Gesamtleistung zu steigern (Ma et al., 2023). Jeder Entscheidungsbaum wird dabei auf einem zufällig gezogenen Teilmengen-Datensatz und einer zufälligen Untermenge der Merkmale trainiert (Bootstrap Aggregation). Durch diese Zufallskomponenten wird das Risiko des Overfittings einzelner Bäume verringert und die Generalisierungsfähigkeit des Gesamtmodells verbessert. Wichtige Hyperparameter eines Random Forests sind u.a. die Stichprobenrate, die Anzahl der Bäume sowie die Tiefe der Bäume. Für Klassifikationsaufgaben wird die finale Entscheidung per Mehrheitsabstimmung der Bäume getroffen. Ma et al. (2023) evaluierten einen Random-Forest-basierten Ansatz zur DDoS-Erkennung in einer SDN-Umgebung unter Verwendung des modernen CIC-DDoS2019-Datensatzes. Der Datensatz wurde vorab umfangreich aufbereitet.

Es erfolgten Datentypkonvertierungen, Merkmalscodierung, das Auffüllen fehlender Werte, das Entfernen redundanter Daten, eine Datenbalancierung sowie eine Normalisierung der Werte. Anschließend wurde der Datensatz in 67% Trainings- und 33% Testdaten aufgeteilt. Zur Merkmalsauswahl kam ein kombinierter Algorithmus zum Einsatz, der fünf Methoden vereint: Varianz, Mutual Information, Backward Elimination, Lasso (L1) und Random Forest. Ziel war die Echtzeiterkennung von DDoS-Angriffen in SDN. Die erzielten Ergebnisse waren sehr hoch. Die Genauigkeit, Präzision, Rückruf und F1-Score lagen jeweils bei 99% (Ma et al., 2023).

Zum Vergleich wurde das Gradient-Boosting-Modell herangezogen. Gradient Boosting (GB) ist ein weiteres Ensemble-Verfahren, bei dem Entscheidungsbäume sequentiell trainiert werden. Jeder neue Baum korrigiert die Fehler des vorherigen. Aldualiji et al. (2022) testeten ein GB-Modell auf den Datensätzen CIC-IDS2017 und CIC-DDoS2019. Nach entsprechender Vorverarbeitung (Umwandlung aller Daten in numerische Formate) wurden die wichtigsten Merkmale mittels Mutual Information und Random-Forest-Feature-Importance ausgewählt. GB erreichte ähnlich hervorragende Ergebnisse wie Random Forest (Genauigkeit = 99% auf beiden Datenansätzen). Bei genauer Betrachtung zeigten sich lediglich minimale Unterschiede in der Fehlklassifikationsrate. Auf CIC-IDS2017 traten 4 Fehlklassifikationen auf, auf CICDDoS2019 nur 2. Random Forest produzierte in einem Szenario mit 19 ausgewählten Merkmalen 16 Fehlklassifikationen (davon 11 FP und 5 FN), erzielte aber dennoch die gleichen Werte für Genauigkeit, Präzision, Rückruf und F1 (jeweils 99%). Insgesamt weisen sowohl Random Forest als auch GB eine sehr hohe Genauigkeit bei der DDoS-Erkennung auf. Random Forest überzeugt durch robuste Generalisierung dank der Kombination vieler Bäume, während GB durch gezielte Fehlerreduktion und Hyperparametersteuerung ebenfalls gute Ergebnisse liefert (Alduailij et al., 2022).

Moderne DL-Ansätze können komplexe Muster direkt aus Rohdaten lernen und erreichen dadurch oft eine noch höhere Erkennungsleistung. Ein beliebtes Beispiel sind CNNs, die speziell für die Klassifikation komplexer Muster (ursprünglich Bilder) entwickelt wurden. CNNs lernen hierarchische Merkmale aus den Eingabedaten und haben sich auch für die Erkennung von DDoS-Angriffen als geeignet erwiesen (Najar und Naik, 2024; Shaaban et al., 2019). Im Kontext von SDN nutzen CNN-Modelle die zentralisierte Steuerung und die Transparenz des Netzwerkverkehrs, um Angriffe effektiv zu erkennen und abzuwehren (Najar und Naik, 2024).

Eine typische CNN-Architektur besteht aus einer Eingabeschicht, gefolgt von mehreren abwechselnden Faltungs- und Pooling-Schichten zur Merkmalextraktion, anschließend vollständig verbundenen Schichten sowie einer Ausgabeschicht. Die Netztiefe hängt von der Anzahl der verwendeten Schichten ab (Shaaban et al., 2019).

Shaaban et al. (2019) entwickelten ein CNN-Modell, das auf zwei unterschiedliche trainiert wurde. Das erste ist der in einem Mobile-Cloud-Computing (MCC)-Testnetzwerk erzeugte Datensatz und das zweite ist der bekannte NSL-KDD-Datensatz. Die Netzwerkdaten wurden in Matrizen transformiert und

als Eingabe für das CNN genutzt, das charakteristische Muster mithilfe von Faltungs- und Aktivierungsschichten lernte. Das Modell erzielt auf beiden Datensätzen eine Genauigkeit von rund 99%. Damit übertraf es die in derselben Studie getesteten klassischen Modelle deutlich. Decision Trees erzielten 94% bzw. 92% Genauigkeit und KNN 93% bzw. 92% (Shaaban et al., 2019).

Ein weiteres CNN-basiertes Verfahren wurde von Haider et al. (2020) vorgeschlagen. Diese Arbeit entwickelte ein Deep-CNN Ensemble-Framework zur Angriffserkennung in SDN. Dabei wurden zwei ähnliche CNN-Modelle parallel trainiert und ihre Ausgaben zu einer Gesamtentscheidung kombiniert. Es wurde auf dem CIC-IDS2017-Datensatz mit rund 140.000 Datenflüssen trainiert. Das Ensemble erzielte eine Genauigkeit von 99,45% und übertraf damit andere DL-Ansätze wie RNNs oder LSTMs in derselben Untersuchung. Durch die Ensemble-Struktur benötigte das Training zwar mehr Zeit, jedoch war die Klassifikation effizient und verbrauchte vergleichsweise geringe CPU-Ressourcen (Haider et al., 2020).

Wie bereits in Kapitel 2.4.2 kurz erwähnt, entwickelten Akgun et al. (2022) ein CNN-basiertes IDS, das auf dem CIC-DDoS2019-Datensatz trainiert wurde und sehr hohe Erkennungsraten erzielte. Das Modell nutzte eindimensionale Faltungsschichten in Kombination mit Inception-ähnlichen Blöcken zur Merkmalsextraktion. Vorverarbeitungsschritte wie die Eliminierung irrelevanter Merkmale, Datenbalancierung und Normalisierung trugen entscheidend zur Leistung bei. Das System erzielte eine Genauigkeit von 99,99% bei binärer Klassifikation sowie 99,30% bei Multiklassenklassifikation und erwies sich auch in Echtzeit als leistungsfähig (Akgun et al., 2022).

Neben den klassischen CNN-Architekturen gibt es tiefergehende Netzwerke, die speziell entwickelt wurden, um noch komplexere Muster zu lernen. Ein bekanntes Beispiel ist das Residual Network (ResNet), das als Erweiterung des CNN gilt und mittels Residual Blocks den Trainingseffekt in sehr tiefen Netzen verbessert und somit einige Schwächen herkömmlicher CNNs überwindet. Hussain et al. (2020) nutzten ResNet18 mit 18 Schichten zur Erkennung von DDoS-Angriffen und trainierten ihr Modell nach dem CIC-DDoS2019-Datensatz. Da Netzwerkdaten nicht bildhaft vorliegen, wurden sie in künstliche Bilder überführt, indem Merkmalswerte normalisiert und zu RGB-Bildern zusammengesetzt wurden. Mit dieser Methode erreichte das ResNet18-Modell 99,99% Genauigkeit bei binärer Klassifikation und 97% bei Mehrklassenklassifikation, während in der Testphase die Genauigkeit 87% erzielte (Hussain et al., 2020).

Bazzi et al. (2020) setzten hingegen ein tieferes Modell, ResNet50 ein. Dafür wurden Netzwerkverkehrsdaten in zweidimensionale Graustufenbilder transformiert und anschließend zum Training genutzt. Zum Einsatz kamen sowohl ein selbsterstellter SYN-Flood-Datensatz als auch der öffentliche CIC-DDoS2019-Datensatz. Das Training über mehrere Epochen zeigte ebenfalls sehr hohe Ergebnisse mit einer Genauigkeit von 99,99% auf dem eigenen Datensatz und 99,56% auf CIC-DDoS2019. Die leicht geringere Genauigkeit auf dem externen Datensatz deutet auf die Herausforderung hin, Modelle über unterschiedliche Netzwerkbedingungen hinweg zu generalisieren (Bazzi et al., 2020).

Vergleichend lässt sich feststellen, dass beide Ansätze, Hussain et al. (2020) mit ResNet18 und Bazzi et al. (2020) mit ResNet50 die Umwandlung von Netzwerkdaten in Bildform nutzen und sehr hohe Genauigkeiten erzielen konnten. Während Hussain et al. die Leistungsfähigkeit einer kompakteren Architektur aufzeigten, verdeutlichten die Ergebnisse von Bazzi et al., dass tiefere Netze zwar ähnlich hohe Genauigkeiten erzielen, jedoch stärker unter Generalisierungsproblemen leiden.

Während ResNet als Erweiterung von CNNs die Stärken bildbasierter Verfahren nutzt, verfolgen andere Modelle einen ganz anderen Ansatz. Statt Daten in Bilder umzuwandeln, betrachten sie die zeitlichen Abfolgen des Netzwerkverkehrs. Ein Beispiel dafür ist der LSTM-Autoencoder von Wei et al. (2023).

Dieses Modell kombiniert zwei DL-Ansätze. Long Short-Term Memory (LSTM) und Autoencoder (AE). LSTM sind eine spezielle Form von rekurrenten neuronalen Netzen, die besonders gut für Sequenzdaten geeignet sind, da sie die Abhängigkeiten über längere Zeiträume hinweg lernen können. Dadurch lassen sich auch subtile Unterschiede zwischen normalem und bösartigem Datenverkehr erkennen. Autoencoder wiederum sind unüberwachte neuronale Netzwerke, die Eingabedaten komprimieren und anschließend wieder rekonstruieren. Sie werden oft für Dimensionsreduktion und Anomalieerkennung eingesetzt.

Im LSTM-AE-Modell von Wei et al. (2023) bestehen sowohl der Encoder- als auch der Decoder-Teil aus mehreren LSTM-Schichten. Der Encoder verdichtet die Eingabesequenz in eine niedrigdimensionale Darstellung, während der Decoder versucht, die ursprünglichen Daten daraus wiederherzustellen. Normale Daten lassen sich dabei gut rekonstruieren, während Angriffe typischerweise einen größeren Rekonstruktionsfehler verursachen. Dieser Fehler dient als Maß für die Anomalieerkennung. Liegt dieser über einem festgelegten Schwellenwert, wird der Datenpunkt als Angriff eingestuft. Das Modell wurde unüberwacht trainiert, d.h. ausschließlich mit normalem Netzwerkverkehr, und lernte dabei, den Rekonstruktionsfehler zu minimieren. Als Datengrundlage diente der CIC-DDoS2019-Datensatz, der in Trainings-, Validierungs- und Testdaten aufgeteilt wurde. Das Training erfolgte nur mit benignen (legitimen) Traffic-Daten, während das Testset sowohl normale als auch Angriffsdaten enthielt. Das LSTM-AE-Modell erreichte eine sehr hohe Erkennungsrate. Die Gesamtgenauigkeit lag über 99%, in einigen Fällen, bspw. bei LDAP-basierten DDoS-Angriffen wurden für Genauigkeit, Präzision, Rückruf und F1-Score Werte von jeweils 99,96% erzielt. Auch für andere Angriffstypen wie DNS-Amplification (96,08%) oder SNMP-basierte Angriffe (96%) zeigte das Modell überzeugende Ergebnisse (Wei et al., 2023).

In einer ähnlichen Studie setzten Mahmoud et al. (2022) den LSTM-AE auf den älteren NSL-KDD-Datensatz ein. Es wurden zwei Szenarien betrachtet: eine binäre Klassifikation (Normal vs. Angriff) und eine Multiklassifikation mit den Kategorien DoS, Probe, R2L, U2R sowie normal. In beiden Fällen erzielte das Modell sehr hohe Werte mit einer Genauigkeit, Präzision, Rückruf und F1-Score von jeweils rund 98-99%. Besonders hervorzuheben ist der AUC-Wert (Area Under the ROC Curve) von 0,999%,

der auf eine exzellente Trennschärfe bei gleichzeitig geringer Fehlalarmrate hinweist (Mahmoud et al., 2022).

Alle vorgestellten KI-Modelle zeigen insgesamt hohe Erkennungsgenauigkeiten bei der DDoS-Erkennung, unterscheiden sich jedoch in bestimmten Aspekten. Die klassischen ML-Modelle wie SVM, Random Forest und GB erreichten mit vergleichsweise geringem Trainingsaufwand und moderaten Datenmengen bereits sehr hohe Genauigkeiten von meist 95% und mehr, in den betrachteten Studien oft um 99% und liefern damit zuverlässige Ergebnisse (RT et al., 2014; Ma et al., 2023). Die modernen DL-Ansätze wie CNN oder ResNet erreichten durch tiefe Netzarchitekturen noch höhere Erkennungsraten von teilweise nahezu perfekten Werten (99% und mehr), benötigten jedoch deutlich größere Datensätze und längere Trainingszeiten (Shaaban et al., 2019; Hussain et al., 2020). Autoencoder-Modelle wie LSTM-AE heben sich ab, da sie unüberwacht trainiert werden können und lediglich normalen Datenverkehr benötigen (Wei et al., 2023). Al-Fawa'reh et al. (2024) beschreiben, dass Zero-Day-Angriffe zu identifizieren ein großes Problem für überwachte Verfahren sind und unüberwachte Ansätze wie Autoencoder Vorteile haben.

Darüber hinaus unterscheiden sich die Ansätze auch in praktischen Eigenschaften. Klassische ML-Modelle arbeiten meist mit tabellarischen Merkmalsvektoren, lassen sich effizient trainieren und eignen sich daher gut für ressourcenschonende Echtzeitanwendungen (Ma et al., 2023). DL-Modelle nutzen dagegen komplexere Datenrepräsentationen wie Bildmatrizen (CNN/ResNet) oder Sequenzdaten (LSTM) und erreichen dadurch höhere Genauigkeit, sind jedoch rechenintensiver (Hussain et al., 2020). Besonders in SDN-Umgebungen zeigen beide Ansätze Vorteile, da die zentrale Steuerung eine globale Sicht auf den Datenverkehr erlaubt und Angriffe so nahezu in Echtzeit erkannt und abgewehrt werden können (Li et al., 2018).

Insgesamt unterscheiden sich die Modelle nicht nur in ihrer Genauigkeit, sondern auch in ihren Trainingsanforderungen und Einsatzszenarien, sodass die Wahl stark von den verfügbaren Ressourcen und der jeweiligen Netzwerkumgebung abhängt.

3.4 Vor- und Nachteile von KI in der DDoS-Erkennung

Der Einsatz von KI in der Erkennung von DDoS-Angriffen bietet zahlreiche Vorteile, bringt jedoch auch spezifische Herausforderungen mit sich. Während Kapitel 2.5.3 bereits auf die allgemeinen Schwächen und Risiken bei der Integration von KI in Sicherheitsarchitekturen eingegangen ist, sollen im Folgenden insbesondere die Stärken und Grenzen im Kontext der DDoS-Erkennung betrachtet werden.

ML-Modelle erreichen beispielsweise auch mit moderaten Datensätzen Genauigkeiten von über 95%, was ihre Praxistauglichkeit in ressourcenbeschränkten Szenarien unterstreicht (vgl. Kapitel 3.3).

Zudem ermöglichen insbesondere anomaliebasierte IDS, die auf ML-Technologien beruhen, eine größere Anpassungsfähigkeit bei der Erkennung neuer Angriffe. Dadurch lassen sich die langen Entwicklungszeiten signaturbasierter Systeme umgehen, da ML-Modelle aus vorhandenen Daten lernen und Vorhersagen über unbekannte Muster treffen können (Al-Shareeda et al., 2023).

Neben den allgemeinen Stärken von ML zeigen auch Modelle wie Naive Bayes (NB) und Decision Trees spezifische Vorteile, die in der DDoS-Erkennung gezielt eingesetzt werden können. So kann NB neue Ansätze zur DDoS-Erkennung entwickeln und in Verbindung mit IDS eingesetzt werden, um Angriffe vorherzusagen. Das Modell ist zudem leicht zu implementieren, auch mit kleinen Datensätzen effizient trainierbar und weist eine sehr schnelle Verarbeitungsgeschwindigkeit auf. Decision Trees wiederum eignen sich für kostengünstige Umgebungen, da sie eine intuitive Wissensdarstellung bieten, einfach zu implementieren sind und eine hohe Klassifizierungsgenauigkeit erreichen. Nach dem Training sind nur geringe Rechenaufwände erforderlich, was zu einer schnellen Verarbeitung führt. Darüber hinaus sind Flow-Table-Angriffe leichter zu erkennen als Bandbreitenangriffe, was die Erkennung in bestimmten Szenarien zusätzlich erleichtert (Le et al., 2020).

Demgegenüber stehen jedoch verschiedene Nachteile klassischer ML-Ansätze. Zum einen erfordern sie häufig einen höheren menschlichen Aufwand beim Training und der Anwendung (Al-Shareeda et al., 2023). Darüber hinaus sind einzelne Modelle mit spezifischen Schwächen verbunden. Decision Trees reagieren instabil auf bereits geringfügige Änderungen im Trainingsdatensatz. Random Forest weist zwar eine hohe Genauigkeit und Robustheit auf, jedoch auch eine vergleichsweise langsame Trainingsgeschwindigkeit (Le et al., 2020; Ma et al., 2023). Auch SVM benötigt für das Training großer Datensätze erheblichen Rechenaufwand, was seine Anwendungen einschränken kann, bietet aber, wie bereits in Kapitel 3.3 gezeigt, im Gegenzug eine hohe Genauigkeit, geringe Fehlalarmrate und gute Generalisierungsfähigkeit (RT et al., 2014). GB schließlich reduziert Overfitting und erreicht eine hohe Genauigkeit, erfordert jedoch intensives Parameter-Tuning und kann im Vergleich zu Random Forest höhere Fehlklassifikationsraten aufweisen (Alduailij et al., 2022).

Im Gegensatz dazu bieten DL-Modelle verschiedene zusätzliche Vorteile. Sie gelten als besonders leistungsfähig bei großen Datenmengen und erreichen in solchen Szenarien hohe Genauigkeits- und Erkennungsraten (Al-Shareeda et al., 2023). Die in Kapitel 3.3 vorgestellten CNN- und ResNet-Modelle bestätigen diese Stärke, da sie bei der Verarbeitung großer, bildartig aufbereiteter Daten nahezu perfekte Erkennungsraten erreichten.

Ein weiterer Vorteil von DL-Modellen ist das Reduzieren des menschlichen Aufwands, da sie die Merkmalsextraktion und Klassifizierung eigenständig übernehmen. Somit stellen sie ein besonders effizientes Werkzeug dar. Darüber hinaus können sie auch mit unvollständigen Daten arbeiten und sind geeignet, Low-Rate-Angriffe zu erkennen, da sie zeitliche Muster und deren Abhängigkeiten erfassen können. Durch den Einsatz von GPU sind DL-Ansätze bei der Berechnung zudem sehr effizient und ermöglichen eine schnelle Verarbeitung (Mittal et al., 2023).

Dennoch sind auch DL-Modelle mit Nachteilen verbunden. Insbesondere der hohe Rechenaufwand und die Komplexität stellen Herausforderungen dar, die ihre Nutzung in Echtzeitszenarien einschränken können (Mittal et al., 2023). CNN erreichen zwar eine sehr hohe Erkennungsgenauigkeit bei der Identifizierung von DDoS-Angriffen in SDN-Umgebungen und zeigen, wie in Kapitel 3.3 deutlich wurde, eine bessere Erkennungsleistung als klassische ML-Modelle (Najar und Naik, 2024; Hussain et al., 2020), doch ihre Trainings- und Testzeit ist deutlich höher als bei anderen DL-Ansätzen wie LSTM. Zudem arbeiten CNN nicht effizient mit nicht-bildbasierten oder niedrigdimensionalen Datensätzen, da sie ursprünglich für Bildverarbeitungsaufgaben entwickelt wurden. Vorverarbeitungsschritte wie Padding können das Lernverhalten zusätzlich beeinflussen (Najar und Naik, 2024; Hussain et al., 2020). Dies wirkt sich auch auf Anforderungen wie die MTT aus, die nach Somani et al. (2017) entscheidend für die Wirksamkeit eines Abwehrsystems ist. Trotz hoher Genauigkeiten bleibt fraglich, ob DL-Modelle in praktischen Echtzeitszenarien die notwendige Abwehrgeschwindigkeit erreichen können.

Auch LSTM-AE zeigen deutliche Vorteile. Sie sind in der Lage, unbekannte Angriffe zu erkennen, feine Musterunterschiede zwischen gutartigem und bösartigem Verkehr zu identifizieren und die typischen Probleme verschwindender oder explodierender Gradienten bei RNN zu umgehen. Damit stellen sie ein effektives Werkzeug zur DDoS-Erkennung dar (Wei et al., 2023; Mahmoud et al., 2022).

Ihre Leistung hängt jedoch stark von Parametern wie die Lernrate und Zeitfensterlänge ab, die maßgeblich die Konvergenzgeschwindigkeit und Erkennungseffizienz beeinflussen. Ein weiteres Problem stellt der Mangel an umfassenden, öffentlich zugänglichen Datensätzen dar, da Opferorganisationen relevante Informationen häufig nicht preisgeben. Bestehende Datensätze sind daher oft unausgewogen oder nicht ausreichend vielfältig. Darüber hinaus konzentriert sich ein Großteil der Forschung auf die binäre Klassifizierung von Angriffen, während Multiklassenklassifizierungen seltener berücksichtigt werden. Hinzu kommt, dass Modelle zwar mit bekannten Daten sehr hohe Leistungen erzielen, bei Zero-Day-Angriffen jedoch ungenau sind und regelmäßig aktualisiert werden müssen. Schließlich wurden viele DL-Ansätze bisher überwiegend offline evaluiert, sodass ihre Wirksamkeit in realen Netzwerken noch nicht abschließend geklärt ist (Mittal et al., 2023).

Die Betrachtung der Vor- und Nachteile verdeutlicht, dass der Einsatz von KI-Methoden in der DDoS-Erkennung zwar ein hohes Potenzial bietet, jedoch auch mit spezifischen Einschränkungen verbunden ist. Um Aspekte greifbarer zu machen, ist es sinnvoll, die Leistungsfähigkeit der Modelle nicht nur allgemein, sondern im Kontext unterschiedlicher DDoS-Angriffstypen zu untersuchen. Kapitel 4 widmet sich daher dem Vergleich verschiedener Angriffsszenarien und analysiert, wie die vorgestellten KI-Modelle deren Erkennung bewältigen.

4 Vergleich verschiedener DDoS-Angriffstypen und deren Erkennung durch KI

In diesem Kapitel wird der Vergleich der Leistungsfähigkeit der vorgestellten Modelle in den vorherigen Kapiteln in Bezug auf unterschiedliche DDoS-Angriffstypen in den Vordergrund genommen. Ziel hierbei ist es, aufzuzeigen, welche Angriffsmethoden für KI besonders schwer zu erkennen sind, welche Modelle in welchen Szenarien am effektivsten arbeiten und welche Faktoren die Erkennungsleistung beeinflussen. Dabei werden sowohl metrische Kriterien wie Genauigkeit, Erkennungsrate und Fehlalarme als auch qualitative Aspekte wie die Anpassungsfähigkeit und Effizienz berücksichtigt. Die Analyse bildet die Grundlage für die anschließende Diskussion über die Grenzen und Potenziale von KI-gestützter DDoS-Erkennung.

4.1 Kriterien für den Vergleich

Zur Bewertung und zum Vergleich von DDoS-Erkennungssystemen werden verschiedene Evaluationsmetriken herangezogen. Die Grundbegriffe TP, TN, FP, FN und FPR wurden bereits in Kapitel 2.4.2 erläutert. Aufbauend darauf wird im Folgenden die zusätzliche Metrik betrachtet:

- True Negative Rate (TNR): Anteil der korrekt als legitim klassifizierten Pakete (TNR = TN / (TN + FP) (Behal und Kumar, 2017)

Bereits eingeführte Metriken wie Genauigkeit, Präzision, Rückruf und F1-Score werden in diesem Zusammenhang ebenfalls berücksichtigt (vgl. Kapitel 2.4.2). Bei der Bewertung ist ein Kompromiss zwischen FPR und Rückruf zu beachten. Ein niedriger Schwellenwert kann den Rückruf erhöhen, führt jedoch zu mehr Fehlalarmen, während ein hoher Schwellenwert die Fehlalarme reduziert, aber das Risiko erhöht, Angriffe zu übersehen. Das Ziel besteht darin, TP und TN zu maximieren und gleichzeitig FP und FN zu minimieren (Behal und Kumar, 2017).

Weitere Kennzahlen wie der Negative Predictive Value (NPV), die F-Measure Complement (FMC) oder die Klassifizierungsrate (CR) werden in einzelnen Studien genutzt, spielen in der vorliegenden Arbeit jedoch keine zentrale Rolle.

4.2 Vergleichsübersicht: Angriffstypen vs. Modellleistung

In diesem Abschnitt wird eine vergleichende Übersicht der in Kapitel 3.3 vorgestellten Modelle zur DDoS-Erkennung gegeben. Dabei werden sowohl die betrachteten Angriffstypen als auch die Leistungskennzahlen der Modelle analysiert, um Unterschiede in Genauigkeit, Fehlklassifikationen und Erkennungsfähigkeit zwischen klassischen ML- und modernen DL-Ansätzen aufzuzeigen. Der Schwerpunkt liegt auf der Fähigkeit der Modelle, verschiedene Angriffstypen zuverlässig zu erkennen und auf ihre jeweilige Robustheit gegenüber komplexen oder unbekannten Bedrohungen.

Modell	Quelle	Angriffstypen	Accuracy	Precision	Recall	F1-Score
CNN	Najar and Naik, 2024	MSSQL, NETBIOS, UDP, UDPLAG	98,64%	99%	99%	99%
ResNet	Bazzi et al., 2024	SYN Flood DDoS Attacks	96,5%	96,7%	96,5%	96,5%
LSTM-AE	Wei et al., 2021	DNS-, LDAP-, SNMP-Angriffe	99%	99%	99%	99%
SVM	Li et al., 2018	UDP Flood, ICMP Flood, SYN Flood	Nicht angegeben	97%	95%	93%
Random Forest	Ma et al., 2023	Nicht spezifiziert	99%	99%	99%	99%
Gradient Boosting	Alduailij et al., 2022	Brute-Force-Angriffe	99%	99%	99%	99%

Tabelle 1: Übersicht der Modelle zur DDoS-Angriffserkennung mit betrachteten Angriffstypen und Leistungskennzahlen

In Tabelle 1 sind die in Kapitel 3.3 vorgestellten Modelle zur DDoS-Erkennung mit den jeweils betrachteten Angriffstypen und zentralen Leistungsmetriken dargestellt. Die angegebenen Werte repräsentieren die Gesamtergebnisse der Studien und beziehen sich auf die Erkennung aller im Experiment analysierten Angriffe. Da unterschiedliche Datensätze genutzt wurden, ist ein direkter Vergleich der Modelle nur eingeschränkt möglich.

Es wird deutlich, dass sowohl klassische ML-Modelle (z.B. SVM, Random Forest, GB) als auch moderne DL-Ansätze (z.B. CNN, ResNet, LSTM-AE) in der Lage sind, DDoS-Angriffe zuverlässig zu erkennen. Einige Modelle wie das CNN von Naja und Naik (2024) unterscheiden sogar zwischen mehreren spezifischen Angriffstypen. Dazu zählen MSSQL- und NetBIOS-Angriffe, die Protokollschwachstellen ausnutzen, sowie UDP- und UDPLAG-Angriffe, die große Mengen von Paketen versenden und so Netzwerkbandbreiten oder Client-Server-Verbindungen überlasten.

Andere Studien, etwa von Wei et al. (2021), konzentrieren sich auf reflektionsbasierte Angriffe wie DNS, LDAP oder SNMP, bei denen Opfer mit verstärkten Antwortpaketen überflutet werden (Sharafaldin et al., 2019). Alduailij et al. (2022) untersuchten hingegen Brute-Force-Angriffe über Botnets,

welche die Netzwerkgeräte infizieren, um die Systeme des Opfers gezielt zu stören. Random-Forest- und GB-Modelle betrachten DDoS-Angriffe überwiegend als Gesamtheit, ohne zwischen einzelnen Typen zu differenzieren.

Die Ergebnisse in der Tabelle 1 zeigen, dass alle Modelle hohe Leistungswerte erreichen, mit einer Genauigkeit, Präzision, Rückruf und F1-Score von meist über 95%, was die Effektivität der unterschiedlichen Ansätze in der DDoS-Erkennung bestätigt. Ergänzend geben einige Studien auch die Anzahl von FP- und FN-Erkennungen an. So weist GB in der Studie mit dem CIC-DDoS2019-Datensatz jeweils nur einen FP und FN auf (Alduailij et al., 2022), während das optimierte Random-Forest-Modell einen FP und vier FN meldete (Ma et al., 2023). ResNet zeigt 79 FP bei keiner FN-Erkennung (Bazzi et al., 2024), während der LSTM-AE bei DNS-Angriffen etwa 200 FP und 1100 FN sowie bei LDAP-Angriffen 10 FP und 20 FN erreichte. Die hohen FP- und FN-Werte des LSTM-AE bei einzelnen Angriffstypen wie DNS zeigen, dass aggregierte Kennzahlen wie die Genauigkeit und F1 nicht immer Erkennungsleistung im Detail widerspiegeln. Dennoch bleibt das Modell relevant, da es insgesamt sehr hohe Genauigkeiten erreichte, ohne auf große Mengen gelabelter Daten angewiesen zu sein. Insbesondere seine Fähigkeit, zeitliche Muster im Netzwerkverkehr zu erfassen, macht es vielversprechend für die Erkennung neuer oder komplexer Angriffsszenarien (Wei et al., 2021). Für SVM und CNN wurden keine FP/FN-Werte angegeben. Diese Angaben verdeutlichen, dass trotz hoher aggregierter Kennzahlen die Erkennung einzelner Angriffstypen unterschiedlich zuverlässig sein kann.

Die Analyse verdeutlicht, dass Reflektions- und Exploitations-basierte Angriffe unterschiedlich anspruchsvoll für die Modelle sind und dass sowohl die Wahl des Modells als auch die des Datensatzes maßgeblich die Erkennungsleistung bestimmen. Die Übersicht in Tabelle 1 erleichtert damit den Vergleich der Modelle hinsichtlich Effizienz, Flexibilität und praktischer Anwendbarkeit in realen Netzwerkszenarien.

Zusammenfassend zeigt die Übersicht, dass sowohl klassische ML-Modelle wie SVM, Random Forest und GB als auch moderne DL-Ansätze wie CNN, ResNet und LSTM-AE sehr hohe Erkennungsleistungen erzielen. Während ML-Modelle durch geringe Komplexität und schnellere Implementierung punkten, überzeugen DL-Modelle insbesondere durch ihre Robustheit und ihre Fähigkeit, auch komplexe oder bislang unbekannte Angriffsmuster zu erkennen. Damit hängt die Wahl des geeigneten Ansatzes stark vom Anwendungskontext ab, etwa von den verfügbaren Daten, den Rechenressourcen und der geforderten Echtzeitfähigkeit.

4.3 Herausforderungen bei der Erkennung spezifischer Angriffstypen

Wie bereits in Kapitel 2 erläutert, lassen sich DDoS-Angriffe verschiedenen Kategorien wie volumetrischen, protokollbasierten und Application-Layer-Angriffen zuordnen. Aufbauend auf den Ergebnissen

aus Kapitel 4.2, in dem die Modelle zur DDoS-Erkennung hinsichtlich ihrer Leistungsfähigkeit bei verschiedenen Angriffstypen verglichen wurden, sollen im Folgenden die spezifischen Herausforderungen bei der Erkennung dieser Angriffe betrachtet werden.

Ältere Datensätze weisen erhebliche Mängel auf. Häufig fehlen klar kategorisierte Informationen, wodurch eine differenzierte Analyse erschwert wird. Während Angreifer einfache Attacken durchführen können, decken viele Datensätze moderne Angriffsmethoden wie MSSQL-, UDP- oder UDPLAG-Angriffe nicht ab (Najar und Naik, 2024). Auch der CIC-DDoS2019-Datensatz eignet sich nicht zur Identifizierung langsamer Low-Rate-Angriffe.

Reflektionsbasierte Angriffe wie MSSQL, NetBIOS, DNS, LDAP oder SNMP-Amplification stellen besondere Herausforderungen dar. Sie nutzen IP-Spoofing, sodass der Traffic zunächst legitim wirkt, während Reflektoren die Antworten an das Opfer leiten. Durch den Verstärkungseffekt entsteht ein Vielfaches an Datenvolumen, was Netzwerke schnell überlastet. Da hierfür legitime Protokolle und Dienste missbraucht werden, ist eine Filterung schwierig, ohne regulären Verkehr zu beeinträchtigen. Die verteilte Ausführung über Botnets erschwert die Erkennung zusätzlich. Obwohl ML- und DL-Ansätze in simulierten Testumgebungen und mit vorbereiteten Datensätzen hohe Genauigkeiten erreichen, zeigen Studien, dass die Übertragbarkeit auf reale Szenarien eingeschränkt ist. Faktoren wie Botnets, Netzwerkrauschen und Bandbreite weichen oft stark von Trainingsdaten ab (Najafmehr et al., 2023).

Protokollbasierte Angriffe wie SYN-Floods und ICMP-Floods nutzen Schwachstellen in TCP- und ICMP-Protokollen, um Server zu überlasten. Mit der Zunahme internetfähiger Geräte, insbesondere im IoT, treten solche Angriffe häufiger auf. Durch IP-Spoofing sind sie schwer zurückzuverfolgen und klassische Erkennungsmethoden erweisen sich als unzureichend (Bazzi et al., 2024).

SDN-Umgebungen sind besonders anfällig für Flooding-Angriffe. Der zentrale Controller stellt ein kritisches Ziel dar, dessen Überlastung das gesamte Netzwerk lahmlegen kann. Da SDN-Switches Datenverkehr nicht eigenständig prüfen, sondern nach Controller-Regeln weiterleiten, wird die Abwehr zusätzlich erschwert. Angreifer können Traffic so gestalten, dass er normalen Mustern ähnelt, was die Erkennung weiter kompliziert (Li et al., 2018).

Brute-Force-Angriffe wirken ebenfalls auf die Verfügbarkeit von Diensten aus, indem Server mit einer sehr hohen Anzahl an Anfragen überlastet werden. Dies führt dazu, dass legitime Benutzer keinen Zugriff auf den Server mehr haben. Zur Erkennung solcher Angriffe ist eine präzise Abstimmung der Erkennungsparameter notwendig, um Fehlklassifikationen zu reduzieren (Alduailij et al., 2022).

Die Tabelle 1 in Kapitel 4.2 zeigt, dass DL-Modelle wie CNN oder LSTM-AE bei reflektionsbasierten Angriffen zwar hohe Genauigkeiten, Präzision, Rückruf und F1-Score erreichen, jedoch weiterhin Herausforderungen bestehen. Insbesondere die Anzahl an FP- und FN-Erkennungen verdeutlicht, dass die

Erkennung einzelner Angriffstypen unterschiedlich zuverlässig ist. Die komplexe Natur der realen Angriffe, kombiniert mit Amplifikation, verschleierter Angreiferidentität und verteiltem Traffic, macht die DDoS-Erkennung trotz leistungsfähiger Modelle anspruchsvoll.

Zusammenfassend verdeutlicht Kapitel 4.3, dass die Erkennung spezifischer Angriffstypen besondere Schwierigkeiten mit sich bringt. Reflektions- und Amplifikationsangriffe sind schwer zu unterscheiden, da sie legitime Protokolle ausnutzen (Najafmehr et al., 2023). Protokollbasierte Angriffe wie SYN- oder ICMP-Floods treten vermehrt in IoT-Umgebungen auf (Bazzi et al., 2024), während SDN-Architekturen durch ihre zentrale Steuerung besonders verwundbar sind (Li et al., 2018). Brute-Force-Angriffe erfordern wiederum eine präzise Parametisierung, um Fehlalarme zu vermeiden (Alduailij et al., 2022). Diese Beispiele zeigen, dass die Grenzen nicht nur in den Modellen selbst liegen, sondern auch in den Eigenschaften der Angriffe und der Datenbasis, die zur Erkennung herangezogen werden.

4.4 Diskussion: Grenzen und Potenziale der KI bei der DDoS-Erkennung

Die bisherigen Ergebnisse verdeutlichen, dass KI-gestützte Verfahren grundsätzlich eine hohe Leistungsfähigkeit bei der DDoS-Erkennung aufweisen, jedoch durch mehrere Faktoren eingeschränkt werden. Neben den bereits beschriebenen Herausforderungen wie hohem Rechenaufwand, mangelnder Generalisierbarkeit auf reale Netzwerke und der Abhängigkeit von Trainingsdaten ergeben sich weitere Diskussionspunkte.

Ein zentralisierter Aspekt ist die Qualität und Standardisierung der Datensätze. Unvollständige oder unsauber kategorisierte Daten führen zu unzuverlässigen Modellen. Wie bereits in Kapitel 4.2 deutlich wurde, erschwert zudem die Nutzung unterschiedlicher Datensätze den Studien die Vergleichbarkeit der Ergebnisse. Dadurch lassen sich kaum allgemeingültige Aussagen zur Leistungsfähigkeit einzelner Modelle ableiten (Najar und Naik, 2024). Auch moderne Datensätze wie CIC-DDoS2019 bilden nicht alle Angriffstypen ab, insbesondere langsame Low-Rate-Angriffe (Najafmehr et al., 2023). Dies zeigt, dass die Wirksamkeit von ML- und DL-Ansätzen nicht allein von der Modellarchitektur abhängt, sondern ebenso von der Qualität und Repräsentativität der zugrunde liegenden Daten.

Faktoren wie Botnets, Netzwerkrauschen, Bandbreite und heterogene Traffic-Muster führen zudem dazu, dass Modelle in realen Umgebungen oft weniger zuverlässig arbeiten als in simulierten Testumgebungen (Najafmehr et al., 2023). Hinzu kommt, dass moderne DL-Modelle erheblichen Ressourcenbedarf aufweisen, da sie enorme Rechenleistung und Speicher benötigen, um in Echtzeit eingesetzt zu werden (Mittal et al., 2023; Bazzi et al., 2024).

Darüber hinaus stellt die Black-Box-Natur vieler DL-Modelle eine Herausforderung dar. Sheed Isael (2025) merkt an, dass fehlende Nachvollziehbarkeit von Entscheidungsprozesse zu einem Mangel an Transparenz führt, was in sicherheitskritischen Bereichen das Vertrauen in KI-basierter Systeme einschränken kann. Des Weiteren hat Sheed Isael zu bemerken, dass KI-Systeme in der Cloud-Sicherheit

oft auf umfangreicher Datenerfassung und -analyse basieren. Diese wirft Bedenken hinsichtlich der Verletzung der Privatsphäre auf. Das Risiko, dass sensible Daten von KI-Systemen oder böswilligen Akteuren unbefugt abgerufen oder missbraucht werden, wird erhöht (von Faber und Kohler, 2019; Pohlmann, 2025).

Hybridansätze bieten eine Möglichkeit, die Grenzen einzelner Modelle zu überwinden. Ein Beispiel hierfür ist das in Kapitel 2.4.2 vorgestellte SVM-SOM-Modell. Während SVM bekannte Muster zuverlässig klassifiziert, kann die SOM-Komponente unbekannte Muster im Netzwerkverkehr identifizieren. In Kombination erreichte das Modell in der Studie von Deepa et al. (2018) sowohl eine höhere Genauigkeit als auch eine geringere Fehlalarmrate als die Einzelmodelle. Damit zeigt sich, dass durch geschickte Modellintegration die Schwächen einzelner Verfahren kompensiert werden können.

Darüber hinaus spielt die Merkmalsauswahl (Feature Selection) eine entscheidende Rolle. Wie in Kapitel 3.2 erläutert, verbessert die Reduktion auf besonders aussagekräftige Merkmale sowohl die Genauigkeit als auch die Effizienz der Modelle, da weniger Daten verarbeitet werden müssen. Gleichzeitig besteht jedoch das Risiko, dass durch eine unzureichende Auswahl relevante Informationen verloren gehen. Dies kann dazu führen, dass Angriffe nicht erkannt oder falsch klassifiziert werden (Saha et al., 2023; Ma et al., 2023). Damit wird deutlich, dass die Qualität der Vorverarbeitung und die Wahl geeigneter Merkmale genauso wichtig sind wie die Modellarchitektur selbst.

Trotz dieser Einschränkungen eröffnen KI-gestützte Systeme erhebliche Potenziale. Adaptive und kontinuierlich lernende Verfahren ermöglichen es, Modelle auch im laufenden Betrieb mit neuen Daten zu aktualisieren und sie so an veränderte Angriffsmuster anzupassen. In Verbindung mit hybriden Ansätzen könnten so robustere und praxisnähere Lösungen entstehen (Lu und Ding, 2021; Pohlmann, 2025).

Zusammenfassend zeigt sich, dass die Leistungsfähigkeit KI-gestützter DDoS-Erkennungssysteme nicht allein von der Modellarchitektur abhängt. Datenqualität, Interpretierbarkeit, Ressourcenbedarf sowie die Integration verschiedener Verfahren und Merkmalsauswahlmethoden spielen ebenso eine große Rolle. Zukünftige Entwicklungen sollten daher auf der Kombination leistungsfähiger Modelle mit optimierten Datenstrategien und adaptiven Mechanismen basieren, um den Anforderungen realer Netzwerkszenarien gerecht zu werden.

5 Ausblick und zukünftige Entwicklungen

In Kapitel 4 wurde ersichtlich, dass KI-gestützte Ansätze ein großes Potenzial für die DDoS-Erkennung bieten. Jedoch bringen sie auch gleichzeitig bestimmte Einschränkungen mit sich. Insbesondere ihre Abhängigkeit von Trainingsdatensätzen, die Komplexität moderner Angriffsmuster und die begrenzte Transparenz vieler DL-Modelle stellen Herausforderungen für die praktische Umsetzung dar.

In Kapitel 5 werden mögliche Weiterentwicklungen und Strategien vorgestellt, die dazu beitragen können, die Erkennungseffizienz zu steigern, die Robustheit gegenüber Angriffstypen zu erhöhen und die Anwendbarkeit von KI-Systemen in realen Netzwerken zu verbessern.

5.1 Trends in KI und Cybersicherheit

Aufbauend auf den Ergebnissen aus Kapitel 4, in dem die Leistungsfähigkeit verschiedener KI-Modelle bei der DDoS-Erkennung sowie deren Grenzen wie hoher Rechenaufwand, eingeschränkte Generalisierung und das Black-Box-Problem aufgezeigt wurden, zeigen aktuelle Trends, wie neue Methoden diese Schwächen adressieren. Ziel ist es, die Erkennungsleistung, Anpassungsfähigkeit und Sicherheit von KI-Systemen weiter zu erhöhen.

Ein zentraler Trend ist die Nutzung von Hardware Performance Counters (HPC). HPCs überwachen detaillierte Leistungsinformationen von Prozessen und liefern Echtzeiteinblicke, die eine effektive und effiziente Erkennung von DDoS-Angriffen, insbesondere in industriellen Systemen ermöglichen. ML-Ansätze werden zudem in Kombination mit SDN in SDN-Controllern implementiert, um zentralisierte Erkennungsdienste zu schaffen. Diese Ansätze nutzen die Netzwerkressourcen effizient und erzielen eine hohe Genauigkeitsrate, etwa 98,2% mit Naive Bayes oder 97,1% mit SVM und Decision Tree (Oun et al., 2025).

Ein weiterer wichtiger Trend ist das Federated Learning (FL-DAD), welches den Datenschutz in der DDoS-Erkennung verbessert. Hierbei erkennen IoT-Geräte oder -Cluster Bedrohungen unabhängig voneinander und können nahezu in Echtzeit auf sich ändernde Angriffsmuster reagieren. Gleichzeitig bleibt die Verarbeitung dezentral, was Vorteile hinsichtlich Datenschutzes und Skalierbarkeit bietet (Oun et al., 2025).

Für die zukünftige Entwicklung der DDoS-Erkennung wird besonders auf die Verbesserung der Generalisierung und Skalierbarkeit von KI-Modellen geachtet. Modelle sollen anhand vielfältiger Datensätze trainiert werden, um ein breiteres Spektrum an Angriffen zuverlässig zu erkennen und realistischere Szenarien abzudecken. Zudem ist die Überprüfung in realen Umgebungen entscheidend, um sicherzustellen, dass die Systeme nicht nur unter Laborbedingungen, sondern auch in der Praxis effektiv arbeiten (Oun et al., 2025; Diop et al., 2025). Ergänzend dazu besteht Forschungsbedarf bei der Balance zwischen Genauigkeit und Fehlalarmen, um eine hohe Sensitivität bei gleichzeitig geringen FP zu erreichen. Ein Beispiel hierfür ist das in Kapitel 3.3 beschriebene LSTM-AE-Modell von Wei et al. (2023), das trotz sehr hoher Genauigkeit aggregiert betrachtet bei einzelnen Angriffstypen wie DNS eine große Zahl FP- und FN-Erkennungen aufwies. Dies verdeutlicht, dass hohe Genauigkeit allein nicht ausreicht und eine verbesserte Balance zwischen Sensitivität und Fehlalarmen notwendig bleibt.

Ein weiterer Fokus liegt auf kontinuierlicher Echtzeitüberwachung. Hier sollen Algorithmen effizienter gestaltet werden, indem leichtgewichtige Modelle entwickelt werden, die den Rechenaufwand reduzieren und gleichzeitig schnelle Reaktionen auf Bedrohungen ermöglichen (Oun et al., 2025).

Besonders das in Kapitel 4 hervorgehobene Black-Box-Problem moderner DL-Modelle erfordert Lösungen, die über reine Leistungskennzahlen hinausgehen. Explainable AI (XAI) gewinnt daher zunehmend an Bedeutung. XAI-Methoden ermöglichen es, Entscheidungen von ML- und DL-Modellen transparent und nachvollziehbar zu machen. In sicherheitskritischen Kontexten ist dies entscheidend, um Vertrauen zu schaffen, Fehlalarme gezielt zu reduzieren und Compliance-Anforderungen zu erfüllen. Neben Transparenz spielt auch Datenschutz eine zentrale Rolle. KI-Systeme in der Cloud-Sicherheit verarbeiten große Mengen sensibler Daten, deren Schutz gewährleistet werden muss. Technologien wie Blockchain können hier einen zusätzlichen Sicherheitsgewinn bringen, während XAI die Grundlage für vertrauenswürdige und überprüfbare KI-basierte Entscheidungen schafft (Oun et al., 2025).

Abschließend ist die kontinuierliche Anpassung an sich entwickelnde Bedrohungen entscheidend. DDoS-Angriffe entwickeln sich parallel zu den Fortschritten in der KI weiter, sodass verbesserte ML- und DL-Modelle rechtzeitig implementiert werden müssen, um Netzwerke vor Angriffen und Datendiebstahl zu schützen. Zudem können kollaborative Erkennungs- und Präventionsmethoden dazu beitragen, Angriffe frühzeitig an unterschiedlichen Orten zu identifizieren und gleichzeitig die Sicherheit und Vertraulichkeit von Kundendaten zu gewährleisten (Ahmadi, 2024).

6 Fazit

In diesem abschließenden Kapitel werden die zentralen Ergebnisse der Arbeit zusammengefasst und eine Antwort auf die Forschungsfrage gegeben. Ziel ist es, die wesentlichen Erkenntnisse komprimiert darzustellen, die gewonnenen Einsichten kritisch zu reflektieren und ihre Bedeutung für Forschung und Praxis einzuordnen.

Neben den in dieser Arbeit untersuchten ML- und DL-Ansätzen prägen auch neue Entwicklungen wie LLM und GenAI zunehmend die Diskussion rund um den Einsatz von KI in der IT-Sicherheit. Während diese Systeme derzeit vor allem in der Text- und Bildgenerierung genutzt werden, könnten sie in Zukunft auch in der Cybersicherheit an Bedeutung gewinnen, etwa zur automatisierten Angriffssimulation oder zur Unterstützung von Threat-Intelligence-Systemen. Auch die Vision einer starken KI (AGI) verdeutlicht, wie weitreichend künftige Entwicklungen sein könnten, wenngleich sie für die konkrete DDoS-Erkennung aktuell keine direkte Relevanz besitzen.

6.1 Zusammenfassung der Ergebnisse

Die Arbeit hat verdeutlicht, dass klassische Abwehrmechanismen den zunehmend komplexen DDoS-Angriffen nur begrenzt standhalten können. Kapitel 2 stellte hierzu die Grundlagen dar. Es beschrieb die typischen Angriffskategorien wie volumetrische, protokollbasierte und Application-Layer-Angriffe und zeigte deren Auswirkungen auf die Verfügbarkeit und Sicherheit von Netzwerken. Zudem wurde deutlich, dass konventionelle Schutzmechanismen wie Firewalls und IDS/IPS in vielen Fällen an ihren Grenzen stoßen.

Aufbauend darauf wurden in Kapitel 3 die Funktionsweisen von ML- und DL-Ansätzen erläutert, Anforderungen an KI-gestützte Erkennungssysteme definiert sowie verschiedene Modelle und Evaluationsmetriken vorgestellt. In Kapitel 4 wurden anschließend die Leistungsfähigkeit dieser Modelle bei der Erkennung unterschiedlicher Angriffstypen analysiert. Dabei zeigte sich, dass sowohl klassische ML-Modelle als auch moderne DL-Ansätze insgesamt hohe Erkennungsraten erzielen. Besonders DL-Modelle wie CNNs und LSTM-AE sind in der Lage, komplexe Muster automatisch zu erfassen und relevante Merkmale direkt aus den Daten zu extrahieren. Dies führte zu höherer Genauigkeit, reduzierten Fehlklassifikationen und einer verbesserten Erkennung unbekannter Angriffstypen. Gleichzeitig verdeutlichte die Analyse spezifische Grenzen. Die Modelle sind stark von der Qualität und Aktualität der Trainingsdaten abhängig, die Erkennung einzelner Angriffstypen kann trotz hoher aggregierter Leistungswerte variieren und Faktoren wie Netzwerkrauschen, Botnets oder heterogene Traffic-Muster in realen Umgebungen können die praktische Wirksamkeit einschränken. Zudem stellen der hohe Rechenaufwand moderner Modelle sowie die eingeschränkte Nachvollziehbarkeit ihrer Entscheidungen durch das Black-Box-Problem zentrale Herausforderungen dar.

Kapitel 5 zeigte schließlich aktuelle Forschungstrends, die dazu beitragen können, diese Grenzen zu überwinden. Dazu zählen u.a. die HPC, FL-DAD, Blockchain-Technologien sowie XAI. Diese Entwicklungen zielen darauf ab, die Generalisierbarkeit und Effizienz zu erhöhen, Datenschutz und Transparenz zu stärken und die Praxistauglichkeit KI-gestützter Systeme zu verbessern.

Insgesamt verdeutlicht die Arbeit, dass KI-basierte Verfahren ein erhebliches Potenzial für die Erkennung und Abwehr von DDoS-Angriffen besitzen. Gleichzeitig hängt ihre Zuverlässigkeit maßgeblich von der Wahl des Modells, der Qualität der Datenbasis und der Anpassungsfähigkeit an reale Netzwerkszenarien ab. Die Ergebnisse liefern damit eine fundierte Grundlage für die abschließende Beantwortung der Forschungsfrage in Kapitel 6.2.

6.2 Beantwortung der Forschungsfrage

Die Forschungsfrage dieser Arbeit lautete: *„Wie effektiv erkennen KI-gestützte Systeme verschiedene DDoS-Angriffstypen und welche Unterschiede bestehen in ihrer Erkennungsleistung?“*

Die Ergebnisse zeigen, dass KI-gestützte Systeme grundsätzlich eine hohe Wirksamkeit bei der DDoS-Erkennung besitzen. Ein Vergleich macht die Unterschiede deutlich. Klassische ML-Modelle wie Random Forest erreichten Genauigkeiten von rund 99% während DL-Modelle wie CNN oder LSTM-AE teils noch höhere Werte (über 98% bis zu 99,99%) erzielten (vgl. Kapitel 4.2). Allerdings verdeutlicht die Analyse, dass hohe Gesamtgenauigkeiten allein nicht ausreichen und die Balance zwischen Sensitivität und Fehlalarmen ein entscheidender Faktor bleibt.

Aufbauend auf den in Kapitel 2 beschriebenen Grundlagen wird deutlich, dass sich die Angriffstypen hinsichtlich des Zieles, Technik und Intensität stark unterscheiden. Während volumetrische Angriffe vor allem durch hohe Datenraten auffallen, zeichnen sich anwendungsbezogene Angriffe wie HTTP-

Floods durch subtilere Muster aus, die schwerer zu erkennen sind. Diese Unterschiede legen die Notwendigkeit spezialisierter Erkennungsansätze nahe.

Wie in Kapitel 4.3 dargestellt, ist die Erkennungsleistung stark vom Angriffstyp abhängig. So erzielten CNN eine Genauigkeit von über 98% bei MSSQL-, NetBIOS- und UDP-Angriffen, während LSTM-AE mit hoher Genauigkeit insbesondere bei DNS- und LDAP-Angriffen überzeugten, jedoch hohe FP- und FN-Raten aufwiesen. Gradient-Boosting-Modelle zeigten ihre Stärken bei Brute-Force-Angriffen mit ähnlich hohen Genauigkeiten, während ResNet-Modelle bei SYN-Floods sehr effektiv waren, aber Schwächen bei Low-Rate-Angriffen zeigten (vgl. Kapitel 4.2). Brute-Force-Angriffe erfordern zudem eine besonders präzise Parametisierung, um Fehlalarme zu vermeiden. Dies macht deutlich, dass sowohl die Wahl des Modells als auch die Art des Angriffs entscheidend für die Zuverlässigkeit der Erkennung sind.

Diese Ergebnisse müssen kritisch im Hinblick auf Datenqualität, Ressourcenbedarf und Interpretierbarkeit betrachtet werden. Wie bereits in Kapitel 4.4 herausgearbeitet, schränken die Abhängigkeit von realitätsnahen Datensätzen, die Black-Box-Natur moderner DL-Modelle und ihr hoher Rechenaufwand den unmittelbaren Praxiseinsatz erheblich ein. Für die IT-Sicherheit bedeutet dies, dass KI-Systeme nur dann verlässlich sind, wenn sie mit hochwertigen Daten trainiert und in bestehenden Infrastrukturen eingebettet werden.

Für die Zukunft zeigen sich mehrere Perspektiven. Neben Weiterentwicklungen wie XAI, FL-DAD oder adaptiven Verfahren rücken auch übergeordnete Trends in den Vordergrund (Oun et al., 2025). LLMs und GenAI könnten künftig für automatisierte Angriffssimulationen oder Threat-Intelligence-Systeme eingesetzt werden. Die Vision einer AGI, wie bereits in Kapitel 2.5.1 diskutiert, unterstreicht zudem die langfristige Bedeutung der KI-Forschung, auch wenn sie für die DDoS-Erkennung aktuell keine direkte Relevanz hat. Langfristig wirft die Diskussion um AGI eine zentrale Frage auf. Könnte eine solche allgemeine KI fähig sein, völlig neue Angriffsarten ohne explizites Training zu erkennen und eigenständig Abwehrstrategien zu entwickeln? Während dies derzeit hypothetisch bleibt, verdeutlicht es, wie weitreichend die Perspektiven für den Einsatz von KI in der Cybersicherheit sind, sowohl als Chance für die Verteidigung als auch als potenzielles Risiko, wenn solche Systeme in falsche Hände geraten (Pohlmann, 2025).

Damit lässt sich die Forschungsfrage dieser Arbeit beantworten. KI-gestützte Systeme erkennen DDoS-Angriffe effektiv. ML-Modelle überzeugen durch ihre Effizienz, DL-Modelle durch höhere Genauigkeit und Robustheit. Die Unterschiede in der Erkennungsleistung hängen jedoch stark vom Modelltyp, den eingesetzten Datensätzen und den spezifischen Angriffstypen ab.

Literaturverzeichnis

Abou El Houda, Z., Nabousli, D. & Kaddoum, G. (2022). Cost-efficient Federated Reinforcement Learning- Based Network Routing for Wireless Networks. *In: 2022 IEEE Future Networks World Forum (FNWF).* IEEE. doi:10.1109/FNFW55208.2022.00050.

Akgun, D., Hizal, S. & Cavusoglu, U. (2022). A new DDoS attacks intrusion detection model based on deep learning for cybersecurity. *Computers & Security*, 118, 102748. doi:10.1016/j.cose.2022.102748.

Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M. & Malik, F. (2022). Machine-Learning-Based DDoS-Attack Detection Using Mutual Information and Random Forest Feature Importance Method. *Symmetry,* 14(6), 1095. doi:10.3390/sym14061095.

Al-Fawa'reh, M., Abu-Khalaf, J., Szewczyk, P. & Kang, J. J. (2023). MalBoT-DRL: Malware Botnet Detection Using Deep Reinforcement Learning in IoT Networks. *IEEE Internet of Things Journal*, 11(6), S. 9610-9629. doi:10.1109/JIOT.2023.3324053.

Ali, M. (2025). Information Security Threats in Cloud Services. A Comprehensive Overview. Bachelor's thesis, Bachelor of Engineering (Information Technology). Verfügbar unter: https://www.theseus.fi/bitstream/handle/10024/893360 [Zugriff am: 26. August 2025].

Al-Shareeda, M. A., Manickam, S. & Saare, A. (2022). DDoS attacks detection using machine Learning and deep learning techniques: analysis and comparison. *Bulletin of Electrical Engineering and Informatics,* 12(2), S. 930-939. doi:10.11591/eei.v12i2.4466.

Ashfaq, M. F., Malik, M., Fatima, U. & Shahzad, M. K. (2022). Classification of IoT based DDoS Attack using Machine Learning Techniques. *In: 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM).* IEEE. doi:10.1109/IMCOM53663.2022.9721740.

Bala, B. & Behal, S. (2024). AI techniques for IoT-based DDoS attack detection: Taxonomies, comprehensive review and research challenges. *Computers & Security*, 52, 100631. doi:10.1016/j.cosrev.2024.100631.

Bazzi, H. S., Nassar, A. H., Haidar, I. M., Haidar, A. M. & Ziad, D. (2024). ResNet-Based Detection of SYN Floods DDoS Attacks. *In: 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*. IEEE. doi:10.1109/IC2PCT60090.2024.10486707.

Behal, S. & Kumar, K. (2017). Detection of DDoS attacks and flash events using novel information theory metrics. *Computer Networks*, 116, S. 96-110. doi:10.1016/j.comnet.2017.02.015.

Bhuyan, M. H., Bhattacharyya, D. K. & Kalita, J. K. (2015). An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection. *Pattern Recognition Letters*, 51, S. 1-7. doi:10.1016/j.patrec.2014.07.019.

Boffey, D. (2025). Pro-Russian hackers claim to have targeted several UK websites. *The Guardian*, 7.Mai. Verfügbar unter: https://www.theguardian.com/technology/2025/may/07 [Zugriff am 25. Juli 2025].

Cao, Y. (2023). Design and Implementation of an Intelligent Machine Learning System Based on Artificial Intelligence Computing. *In: 2023 2nd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI)*. IEEE. doi:10.1109/ICDACAI59742.2023.00141.

Dalal, K. R. (2020). Analysing the Role of Supervised and Unsupervised Machine Learning in IoT. *In: 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*. IEEE. doi:10.1109/ICESC48915.2020.9155761.

Deepa, V., Sudar, K. M. & Deepalakshmi, B. (2018). Detection of DDoS Attack on SDN control plane using Hybrid Machine Learning Techniques. *In: 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT)*. IEEE. doi:10.1109/ICSSIT.2018.8748836.

Garg, U., Kaur, M., Kaushik, M. & Gupta, N. (2021). Detection of DDoS Attacks using Semi-supervised based Machine Learning Approaches. *In: 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST)*. IEEE. doi:10.1109/ICCMST54943.2021.00033.

Gupta, V., Mishra, V. K., Singhal, P. & Kumar, A. (2022). An Overview of Supervised Machine Learning Algorithm. *In: 2022 11th International Conference on System Modelling & Advancement in Research Trends (SMART)*. IEEE. doi: 10.1109/SMART55829.2022.10047618.

Haider, S., Akhunzada, A., Mustafa, I., Bharat, P., Fernandez, A. & Choo, K.-K. R. (2020). A Deep CNN Ensemble Framework for Efficient DDoS Attack Detection in Software Defined Networks. *IEEE Access,* 8, S.53972-53983. doi:10.1109/ACCESS.2020.2976908.

Hassan, M., Metwally, K. & Elshafey, M. A. (2024). ZF-DDoS: An Enhanced Statistical-Based DDoS Detection Approach using Integrated Z-Score and Fast-Entropy Measures. *In: 2024 6th International Conference on Computing and Informatics (ICCI)*. IEEE. doi:10.1109/ICCI61671.2024.10485097.

Hnamte, V., Najar, A. A., Nhung-Nguyen, H., Hussain, J. & Sugali, M. N. (2024). DDoS attack detection and mitigation using deep neural networks in SDN environment. *Computers & Security*, 138, 103661. doi:10.1016/j.cose.2023.103661.

Hussain, F., Abbas, S. G., Husnain, M., Fayyaz, U. U., Shahzad, F. & Shah, G. A. (2020). IoT DDoS and DDoS Attack Detection using ResNet. *In: 2020 IEEE 23rd International Multitopic Conference (INMIC)*. IEEE. doi:10.1109/INMIC50486.2020.9318216.

Isael, S. (2025). AI for Detecting and Mitigating Distributed Denial of Service (DDoS) Attacks in Cloud Networks. *Preprints.org*. doi:10.20944/preprints202503.1833.v1.

James, G., Abraham, C., Dipte, S., Gaat, A. & Siddiqui, A. (2024). Distributed Denial of Service Attack Mitigation Using Reinforcement Learning. *In: 2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*. IEEE. doi:10.1109/ICECSP61809.2024.10698040.

Jia, J. & Wang, W. (2020). Review of reinforcement learning research. *In: 2020 35th Youth Academic Annual Conference of Chinese Association of Automation (YAC)*. IEEE. doi:10.1109/YAC51587.2020.0337653.

Jyoti, N. & Behal, S. (2021). A Meta-evaluation of Machine Learning Techniques for Detection of DDoS Attacks. *In: 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, S. 522-526. Verfügbar unter: https://ieeexplore.ieee.org/document/9441127 [Zugriff am 13. August 2025].

Kitzmann, A. (2022). Künstliche Intelligenz – Vor- und Nachteile. *In: Künstliche Intelligenz*. Springer, Wiesbaden. doi: 10.1007/978-3-658-37700-7_6.

Kolias, C., Kambourakis, G., Stavrou, A. & Voas, J. (2017). DDoS in the IoT. Mirai and Other Botnets. *Computer,* 50(7), S. 80-84. doi:10.1109/MC.2017.201.

Kukreti, S., Modgil, S. K., Gehlot, N. & Kumar V. (2022). DDoS Attack using SYN Flooding: A Case Study. *In: 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)*. S. 323-329. IEEE. doi:10.23919/INDIACom54597.2022.9763108.

Le, D. T., Dao, M. H. & Nguyen, Q. L. T. (2020). Comparison of machine learning algorithms for DDoS attack detection in SDN. *Information and Control Systems*, 3, S. 59-70. doi:10.31799/1684-8853-2020-3-59-70.

Li, D., Yu, C., Zhou, Q. & Yu, J. (2018). Using SVM to Detect DDoS Attack in SDN Network. *IOP Conference Series: Materials Science and Engineering*, 466(1), 012003. doi:10.1088/1757-899X/466/1/012003.

Li, Y. & Lin, F. (2008). Customer segmentation analysis based on SOM clustering. *In: 2008 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE. doi:10.1109/SOLI.2008.4686353.

Lu, W. & Ding, Y. (2021). A Network Malicious Traffic Detection Method Based on Semi-Supervised Deep Learning. *In: 2021 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. IEEE. doi:10.1109/ICSPCC52875.9564717.

Ma, R., Wang, Q., Bu, X. & Chen X. (2023). Real-Time Detection of DDoS-attacks Based on Random Forest in SDN. *Applied Sciences*, 13(13). doi:10.3390/app13137872.

Mahmoud, M., Kasem, M., Abdallah, A. & Kang, H. S. (2022). AE-LSTM: Autoencoder with LSTM-Based Intrusion Detection in IoT. *In: 2022 International Telecommunications Conference (ITC-Egypt)*. IEEE. doi:10.1109/ITC-Egypt55520.2022.9855688.

Mallapragada, P. K., Jin, R., Jain, A. K. & Liu, Y. (2009). SemiBoost: Boosting for Semi-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11), S. 2000-2014. doi:10.1109/TPAMI.2008.235.

Mazhar, N., Salleh, R., Zeeshan, M., Hameed, M. M. & Khan, N. (2021). R-IDPS: Real time SDN based IDPS system for IoT security. *In: 2021 IEEE 18th International Conference on Smart Communities: Improving Quality of Life Using ICT, IoT and AI (HONET).* IEEE. doi:10.1109/HONET53078.2021.9615449.

Mekala, S., Dasari, K. & Katta, D. (2024). DNS DDoS Amplification Attack Using Multi-Layer Perceptron Classification Algorithm. *In: Proceedings of the 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC).* S. 1355-1360. IEEE. doi:10.1109/AIC61668.2024.10730978

Merkel-Kiss, M. & von Garrel, J. (2022). Systematische Literaturanalyse zum KI-Einsatz und KI-basierten Geschäftsmodellen in produzierenden kleinen und mittleren Unternehmen. *Zeitschrift für Arbeitswissenschaft*, 77, S. 453-468. doi:10.1007/s41449-022-00323-9.

Minhas, M. R., Shafi, Q. M., Khan, S. A., Ahmad, T., Ullah, S. & Buriro, A. (2025). F-OSFA: A Fog Level Generalizable Solution for Zero-Day DDOS Attacks Detection. *IEEE Access*, 13, S. 75157-75170. doi:10.1109/ACCESS.2025.3557822.

Mittal, M., Kumar, K. & Behal, S. (2023). Deep learning approaches for detecting DDoS attacks: a systematic review. *Soft Computing*, 27, S. 13039-13075. doi:10.1007/s00500-021-06608-1.

Naim, A., Alshawaf, S. M., Malik, P. K. & Singh, R. (2023). Effective E-Learning Practices by Machine Learning and Artificial Intelligence. *In: 2023 International Conference on Artificial Intelligence and Smart Communication (AISC)*. IEEE. doi:10.1109/AISC56616.2023.10085391.

Naing, S. K. & Thwel, T. T. (2023). A Study of DDoS Attack Classification Using Machine Learning Classifiers. *In: 2023 IEEE Conference on Computer Applications (ICCA).* IEEE. doi:10.1109/ICCA51723.2023.10182146

Najafimehr, M., Zarifzadeh, S. & Mostafavi, S. (2023). DDoS attacks and machine-learning-based detection methods: A survey and taxonomy. *Engineering Reports*, 5(12), S. 1-29. doi:10.1002/eng2.12697.

Najar, A. A. & Naik, S. M. (2024). Cyber-Secure SDN: A CNN-Based Approach for Efficient Detection and Mitigation of DDoS attacks. *Computers & Security*, 139, 103716. doi:10.1016/j.cose.2024.103716.

Nogueira, M., Santos, A.A. & Moura, J.M.F. (2017). Early Signals from Volumetric DDoS Attacks: An Empirical Study. *arXiv preprint arXiv:1609.09560*. doi: 10.48550/arXiv.1609.09560.

Oktivasari, P., Zain, A., R., Kurniawan, A., Murad, F. A. & Anshor, M. F. (2022). Analysis of Effectiveness of Iptables on Web Server from Slowloris Attack. *In: 2022 5th International Conference of Computer and Informatics Engineering (IC2IE)*. S. 215-219. IEEE. doi:10.1109/IC2IE56416.2022.9970143.

Padhiar, S., Patel, H. & Chocha, D. (2025). Detecting and Preventing DoS/DDoS Attacks through Monitoring and Limiting Malicious Traffic. *In: 2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0.* IEEE. doi:10.1109/OTCON65728.2025.11071154.

Pathak, A. K., Dinker, G. A. & Sharma, S. K. (2025). A Taxonomy of DDoS Attacks on Cloud Computing Environment and Approaches to Mitigation. *In: 2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI)*. IEEE. doi:10.1109/IC3ECSBHI63591.2025.10991034.

Patil, R. Y. & Ragha, L. (2011). A rate limiting mechanism for defending against flooding based distributed denial of service attack. *In: 2011 World Congress on Information and Communication Technologies.* IEEE. doi:10.1109/WICT.2011.6141240.

Pohlmann, N. (2025). IT-Sicherheit und Künstliche Intelligenz. *Datenschutz und Datensicherheit,* 49, S. 5-10. doi:10.1007/s11623-024-2030-y.

Raja, T. V., Ezziane, Z., He, J., Ma, X. & Kazaure, A. W.-Z. (2022). Detection of DDoS Attack on Smart Home Infrastructure Using Artificial Intelligence Models. *In: 2022 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*. IEEE. doi:10.1109/CyberC55534.2022.00014.

Raj, R. & Kang, S. S. (2022). Mitigating DDoS Attack using Machine Learning Approach in SDN. *In: 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. IEEE. doi:10.1109/ICAC3N56670.2022.10074307.

Rincy, T. N. & Gupta, R. (2020). A Survey on Machine Learning Approaches and Its Techniques. *In: 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE. doi:10.1109/SCEECS48394.2020.190.

Roopak, M., Tian, G. Y. & Chambers, J. (2020). An Intrusion Detection System Against DDoS Attacks in IoT Networks. *In 2020 10th Annual Computing and Communication Workshop and Conference (CCWC).* IEEE. doi:10.1109/CCWC47524.2020.9031206.

RT, K., Selvi, S. T. & Govindarajan, K. (2014*). DDoS detection and analysis in SDN-based environment using support vector machine classifier. In: 2014 Sixth International Conference on Advanced Computing (ICoAC).* IEEE. doi:10.1109/ICoAC.2014.7229711.

Rudman, L. & Irwin, B. (2015). Characterization and analysis of NTP amplification based DDoS attacks. *In: 2015 Information Security for South Africa (ISSA).* IEEE. doi:10.1109/ISSA.2015.7335069.

Saha, S., Priyoti, A. T., Sharma, A. & Haque, A. (2022). Towards an Optimal Feature Selection Method for AI-Based DDoS Detection System. *In: 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC).* IEEE. doi:10.1109/CCNC49033.2022.9700569.

Salem, A. H., Azzam, S. M., Emam, O. E. & Abohany, A. A. (2024). Advancing cybersecurity: a comprehensive review of AI-driven detection techniques. *Journal of Big Data,* 11, 105. doi:10.1186/s40537-024-00957-y.

Shaaban, A. R., Abd-Elwanis, E. & Hussein, M. (2019). DDoS attack detection and classification via Convolutional Neural Network (CNN). *In: 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS).* IEEE. doi:10.1109/ICICIS46948.2019.9014826.

Shaaban, A. R., Ad-Elwanis, E. & Hussein, M. (2019). TCP and HTTP Flood DDoS Attack Analysis and Detection for space ground Network. *In: 2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*. IEEE. doi:10.1109/ICVES.2019.8906302.

Shaharkar, B., Mittal, T., Sunkara, K. C., Shukla, V., Desai, I. & Mookherjee, U. (2024). A Multi-Agent Deep Reinforcement Learning Framework for Detecting and Mitigation DDoS Attacks in IoT Networks. *In: 2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC).* IEEE. doi:10.1109/ICMNWC63764.2024.10872370.

Sharafaldin, I., Lashkari, A. H., Hakak, S. & Ghorbani, A. A. (2019). Developing Realistic Distributed Denial of Service (DDoS) Attack Dataset and Taxonomy. *In: 2019 International Carnahan Conference on Security Technology (ICCST).* IEEE. doi:10.1109/CCST.2019.8888419.

Singh, A. & Juneja, D. (2010). Agent Based Preventive Measure for UDP Flood Attack in DDoS Attacks. *International Journal of Engineering Science and Technology*, 2(8), S. 3405-3411. Verfügbar unter: https://www.researchgate.net/publication/50315626 [Zugriff am 30. Juli 2025].

Siriyapuraju, S. J., Gowri, V. S., Balla, S., Vanika, M. K. & Gandhi, A. (2023). DoS and DDoS attack detection using mathematical and entropy methods. *In: 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS).* IEEE. doi: 10.1109/PCEMS58491.2023.10136042.

Somani, G., Gaur, M. S., Sanghi, D., Conti, M., Rajarajan, M. & Buyya, R. (2017). Combating DDoS Attacks in the Cloud: Requirements, Trends, and Future Directions. *IEEE Cloud Computing*, 4(1), S.22-32. doi:10.1109/MCC.2017.14.

Srivastava, A., Tiwari, S., Kumar, D. & Garg, N. (2024). Finding of DDoS Attack in IoT-Based Networks Using Ensemble Technique. *In: 2024 International Conference on Intelligent Systems for Cybersecurity (ISCS).* IEEE. doi:10.1109/ISCS61804.2024.10581044.

Treseangrat, K., Kohali, S. S. & Sarrafpour, B. (2015). Analysis of UDP DDoS cyber flood attack and defense mechanisms on Windows Server 2012 and Linux Ubuntu 13. *In: 2015 International Conference on Computer, Information and Telecommunication Systems (CITS)*. IEEE. doi:10.1109/CITS.2015.7297731.

Teja, K., Abhijith, K., Deepak, O. N., Hanish, T. S., Chaitanya, G. K. & Subramanyam, M. M. (2023). Prevention of Attacks and Flow Control of Firewalls. *In: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE. doi:10.1109/ICACCS57279.2023.10112739.

Varre, D. N. M. R. & Bayana, J. (2022). A Secured Botnet Prevention Mechanism for HTTP Flooding Based DDoS Attack. *In: 2022 3rd International Conference for Emerging Technology (INCET)*. S. 1-5. IEEE. doi:10.1109/INCET54531.2022.9824510.

von Farber, E. & Kohler, A. (2019). Die Lücke: Informationssicherheit in Systemen mit künstlicher Intelligenz. *Datenschutz und Datensicherheit*, 43, S. 434-439. doi:10.1007/s11623-019-1139-x.

Wehbi, K., Hong, L., Al-salah, T. & Bhutta, A. A. (2019). A Survey on Machine Learning Based Detection on DDoS Attacks for IoT Systems. *In: 2019 SoutheastCon. IEEE.* doi:10.1109/SoutheastCon42311.2019.9020468.

Wei, Y., Jang-Jaccard, J., Sabrina, F., Xu, W., Camtepe, S. & Dunmore, A. (2023). Reconstruction-based LSTM-Autoencoder for Anomaly-based DDoS Attack Detection over Multivariate Time-Series Data. *arXiv preprint arXiv:2305.09475*. doi:10.48550/arXiv.2305.09475.

Xiang, Y., Li, K. & Zhou, W. (2011). Low-Rate DDoS Attacks Detection and Traceback by Using New Information Metrics. *IEEE Transactions on Information Forensics and Security*, 6(2), S. 669-680. doi:10.1109/TIFS.2011.2107320.

Yuan, X. (2017). An improved Aprori algorithm for mining association rules. *In: AIP Conference Proceedings*, 1820(1), 080005. doi:10.1063/1.4977361.

Zargar, S. T, Joshi, J. & Tipper, D. (2013). A Survey of Defense Mechanisms Against Distributed Denials of Service (DDoS) Flooding Attacks. *IEEE Communications Surveys & Tutorials,* 15(4), S. 2046-2069. doi: 10.1109/SURV.2013.031413.00127.

Zhijun, W., Qing, X., Jingjie, W., Meng, Y. & Liang, L. (2020). Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Networks. *IEEE Access*, 8, S. 17404-17418. doi: 10.1109/ACCESS.2020.2967478.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel:

Vergleich von DDoS-Angriffstypen und deren Erkennung durch KI-gestützte Systeme

selbständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

18. September. 2025

Datum

Unterschrift