



**TECHNISCHE
UNIVERSITÄT
DRESDEN**

Faculty of Computer Science Institute of Systems Architecture, Chair of Privacy and Data Security

Bachelor-Arbeit

Kontextsensitivität für Netzwerk-Sicherheits-Monitoring

Tobias Köllner

Geboren am: 25. Dezember 1999 in Dresden

Matrikelnummer: 4803900

Immatrikulationsjahr: 2018

zur Erlangung des akademischen Grades

Bachelor of Science (B.Sc)

Betreuer

M.Sc. Sebastian Rehms

Betreuender Hochschullehrer

Dr. Stefan Köpsell



AUFGABENSTELLUNG FÜR DIE BACHELORARBEIT

Name, Vorname des Studenten: Köllner, Tobias
Immatrikulationsnummer: 4803900
Studiengang: Bachelor Informatik
E-Mail: tobias.koellner@mailbox.tu-dresden.de

Thema: Kontextsensitivität für Netzwerk-Sicherheits-Monitoring

Zielstellung:

Zur Erhöhung der Sicherheit von Netzwerken werden häufig Techniken wie Network security monitoring (NSM) sowie intrusion detection (IDS) verwendet. Insbesondere die Menge an falsch-positiven Meldungen stellt hierbei eine große Herausforderung dar. Um dieses Problem sowie die generelle Interpretierbarkeit der Beobachtungsdaten zu verbessern können zusätzliche Informationen gesammelt werden, die man als Netzwerkkontext bezeichnen kann.

Der erste Teil der Arbeit soll untersuchen, wie solche Kontextinformationen sowohl während des Betriebs von NSM-Systemen als auch während der Signatur/Modell-Erstellung im Rahmen von IDS gesammelt und verwendet werden können. Ziele hierbei sind, solche Möglichkeiten auf theoretischer Ebene wie auch von open-source Implementierungen zu sichten und hierbei einen kritischen Kontextbegriff im Vergleich zu etablierten Kontextbegriffen in der Literatur in Bezug auf Netzwerke zu entwickeln.

Aufbauend auf diesen Resultaten kann eine einfache praktische Auswertung verfügbarer Implementierungen mit explizitem Hinblick auf die kontextsensitiven Fähigkeiten durchgeführt werden. Hierfür ist ein Plan für eine minimale Testumgebung sowie das Design von geeignetem Netzwerkverkehr erforderlich.

In der Arbeit sollen schwerpunktmäßig folgende Teilaufgaben bearbeitet werden:

- Überblick und Analyse existierender NSM-Ansätze in Bezug auf Kontext- und Umgebungsbewusstsein. Hierbei sollen mindestens folgende Forschungsfragen bearbeitet werden:
 - Lassen sich generelle Kategorien von Kontext im Rahmen von NSM formulieren?
 - Welche Kontextinformation kann während der Regelgenerierung im Unterschied zu generell verfügbaren Regeln aus öffentlichen Quellen berücksichtigt werden?
 - Welche Techniken wurden vorgeschlagen, um Kontextinformation während des regulären Betriebs zu sammeln und wie werden diese dann genutzt?
 - Inwiefern gibt es Möglichkeiten der Anpassungen an Änderungen der Netzwerkkumgebung, etwa der Topologie oder der genutzten Protokolle?
 - Wie stark trägt Kontextinformation zur Reduktion von falsch-Positiven bei?
- Es soll eine Evaluation der Fähigkeit von Kontextnutzung von verfügbaren open-source-Implementierungen (etwa Plugins in Zeek oder Netzdefinitionen in Suricata) durchgeführt werden.

Verantwortlicher Hochschullehrer:
Institut:
Beginn am: 3.6.2022

Dr. Stefan Köpsell
Systemarchitektur
Einzureichen am: 19.8. 2022

3.6.22, T Köllner
.....
Datum, Unterschrift der/des Studierenden

.....
Unterschrift des betreuenden Hochschullehrers

Selbstständigkeitserklärung

Hiermit versichere ich, dass ich das vorliegende Dokument mit dem Titel *Kontextsensitivität für Netzwerk-Sicherheits-Monitoring* selbstständig und ohne unzulässige Hilfe Dritter verfasst habe. Es wurden keine anderen als die in diesem Dokument angegebenen Hilfsmittel und Quellen benutzt. Die wörtlichen und sinngemäß übernommenen Zitate habe ich als solche kenntlich gemacht. Während der Anfertigung dieses Dokumentes wurde ich nur von folgenden Personen unterstützt:

M.Sc. Sebastian Rehms

Zusätzliche Personen waren an der geistigen Herstellung des vorliegenden Dokumentes nicht beteiligt. Mir ist bekannt, dass die Nichteinhaltung dieser Erklärung zum nachträglichen Entzug des Hochschulabschlusses führen kann.



Tobias Köllner

Zusammenfassung

Um den Betrieb der Computernetzwerke unserer Zeit zu gewährleisten, ist es von äußerster Wichtigkeit, möglichst alle Angriffe und böartigen Netzwerkverkehr zu erkennen, ohne dabei den menschlichen Administrator hinter dem System zu überfordern. Diese Tatsache macht Monitoring und damit Intrusion Detection Systeme zu einem Grundpfeiler einer jeden Strategie der Netzwerksicherheit. Etablierte Implementierungen müssen dabei stetig an sich ändernde äußere Umstände angepasst werden. Einer der zentralen Ansätze, um die Anpassungsfähigkeit gängiger Netzwerkssicherheitskomponenten möglichst ohne menschliches Zutun zu erhöhen oder überhaupt erst zu schaffen, ist Kontextsensitivität, also die Verwendung von Kontext, um dem relevanten Informationen und/oder Dienstleistungen bereitzustellen. In dieser Arbeit wurde diskutiert, wie Kontextinformationen kategorisiert, akquiriert und geformt werden können. Mithilfe des Kontextes wurde die Leistung von Zeek, einem hybriden kontextsensitiven IDS, verbessert und abschließend aufbauend darauf ein Urteil über die einzelnen Kategorien und deren Nützlichkeit zur Erhöhung der Sicherheit abgegeben.

Inhaltsverzeichnis

Zusammenfassung	4
1 Einführung	10
2 Hintergrund	11
2.1 Netzwerk-Sicherheits-Monitoring	11
2.1.1 Intrusion Detection	11
2.1.2 Intrusion Detection System (IDS)	11
3 Voraussetzungen und verwandte Arbeiten	12
3.1 Kontextbegriff	12
3.1.1 Historische Entwicklung	12
3.2 Kontextkategorisierung	13
3.2.1 Übersicht relevanter bereits vorgeschlagener Kategorien	13
4 Design	15
4.1 Notwendigkeit der Erstellung einer Taxonomie	15
4.2 Erstellung einer Taxonomie	15
4.2.1 Mensch	16
4.2.2 Netzwerk	17
4.2.3 Historie der Werte einer Kategorie	19
4.3 Form des Kontextes	20
4.3.1 Historie	21
4.4 Kontextgewinnung	21
4.5 Umwandlung der Taxonomie in IDS-Signaturen	22
4.5.1 Entscheidung für ein IDS	23
4.6 Anspruch an das IDS	24
4.6.1 Auswahl der Kriterien	25
5 Implementierung	26
5.1 Erläuterung der wichtigsten Komponenten	26
5.1.1 Zeek	26
5.1.2 Zeek-Agent	26
5.2 Versuchsaufbau	27
5.2.1 Erzeugung	27
5.2.2 Logging	27

5.3	Skripte	28
5.3.1	Geografische Koordinaten und Ortszeit	28
5.3.2	Verwendete Ports	29
5.3.3	DNS-Auflösung	30
6	Evaluation	31
6.1	Bedeutung von Kontextsensitivität	31
6.2	IDS-Implementierung	32
6.3	Vergleich der Kontextkategorien	32
6.3.1	Verfügbarkeit der Informationen	32
6.3.2	Qualität der Informationen	33
6.3.3	Anwendungsfälle	34
6.3.4	Vertrauenswürdigkeit der Informationsquellen	34
6.4	Leistungsverbesserung	34
7	Zusammenfassung und Ausblick	35

Abbildungsverzeichnis

4.1	Nutzerzentrierte Kategorisierung	17
4.2	Kategorisierung von Kontextinformationen als Bestandteile eines Netzwerkes	19
4.3	Dynamik einer Kontextinformation	20
6.1	Evaluation der Taxonomie	32

Tabellenverzeichnis

4.1	Benötigte Informationen der Historie einer Entität	21
4.2	Akquirierungspunkte für verschiedene Kontextinformationen	22
4.3	Einteilungsansätze für IDS	23

Quelltextverzeichnis

5.1	Konfiguration und Versendung eines Pakets	27
5.2	Generierung einer Log-Datei mit Verbindungsinformationen	28
5.3	Geolokalisierung und Setzen des Grenzwertes	29
5.4	Abfrage und Abgleich der Ports	29
5.5	Überprüfung der Verbindungsziele eines Endgerätes	30

1 Einführung

Seit Computernetzwerke existieren, gibt es auch Menschen, welche sich zu diesen widerrechtlichen Zugang verschaffen wollen. Weder Sicherheitsmaßnahmen noch deren Weiterentwicklung haben diesen Umstand bisher geändert. Lediglich die Netzwerke, welche dabei von Interesse sind, haben sich gewandelt und stellen die IT-Sicherheit dabei immer wieder vor neue Herausforderungen, die zunehmend nicht mehr ausschließlich manuell gelöst werden können. Gleichzeitig verändern sich auch Angreifer und ihre Angriffsarten und die dabei verwendeten Werkzeuge und Techniken entwickeln sich weiter und ermöglichen nach wie vor, in kontrollierte Umgebungen einzudringen, Schäden zu verursachen und dabei trotzdem unentdeckt zu bleiben. Dagegen stehen einem Netzwerkadministrator allerdings eine Vielzahl von Mitteln zur Verfügung. Unter anderem Netzwerk-Sicherheits-Monitoring, dessen erste Stufe beim Umgang mit Angriffen die Erkennung darstellt. Etabliert hat sich dabei vor allem die Verwendung von Intrusion Detection Systemen, welche aber wie jedes System dieses Komplexitätsgrades nie gänzlich fehlerfrei arbeiten. Deshalb ist es letztendlich immer noch nötig, dass die Entscheidungen eines IDS regelmäßig durch einen Menschen überprüft und gegebenenfalls korrigiert werden. Um also die Menge an Meldungen, welche manuell überprüft werden müssen, zu verringern und die Aussagekraft der erzeugten Alarme zu erhöhen und damit den menschlichen Entscheider zu entlasten sowie generell seine Aufgaben besser zu erfüllen, muss ein IDS eine möglichst große Zahl von Szenarien korrekt einschätzen. Dazu werden Informationen benötigt, mit denen die Geschehnisse in einem Netzwerk in einer für das IDS verständlichen Form vollständig abgebildet werden können. In dieser Arbeit soll ergründet werden, auf welche Art welche Informationen automatisch gesammelt und verwertet werden können, um eine Verbesserung der Leistung unter Einbeziehung von zusätzlichem Kontext in die Entscheidungsfindung kontextsensitiver Systeme zu erreichen und welche Kategorien von Kontextinformationen dafür am besten geeignet sind.

2 Hintergrund

2.1 Netzwerk-Sicherheits-Monitoring

Ghafir et al. [9] beschreiben Sicherheits-Monitoring und dessen Aufgaben als eine Reihe von Mechanismen, die ermöglichen, den momentanen Zustand und langfristige Trends eines komplexen Computernetzwerks zu erkennen. Netzwerküberwachung umfasst mehrere Methoden, die gezielt eingesetzt werden, um die Sicherheit und Zuverlässigkeit aufrechtzuerhalten. Dabei reicht das Aufgabenfeld von Hardware, Software, Viren, Spyware und Schwachstellen wie Hintertüren bis zu Sicherheitslücken sowie anderen Aspekten, die die Integrität eines Netzwerks gefährden können. Um proaktiv statt reaktiv vorgehen zu können, muss der Datenverkehr und die Leistung im gesamten Netzwerk überwacht werden, um sicherzustellen zu können, dass keine Sicherheitslücken im Netzwerk auftreten. Im Falle eines Netzwerkfehlers müssen Fehlfunktionen erkannt, isoliert und behoben werden. Bei einem stabilen Netz ist ständige Überwachung, ob eine Bedrohung von innerhalb oder außerhalb vorliegt, notwendig.

2.1.1 Intrusion Detection

Scarfone et al. [20] erklären Angriffserkennung, als Prozess der Überwachung von Ereignissen in einem Computersystem oder Netzwerk und die Analyse dieser Ereignisse auf Anzeichen eines möglichen Zwischenfalls. Gemeint sind damit Verstöße oder unmittelbare Bedrohungen von Sicherheitsrichtlinien, Akzeptanzrichtlinien oder Standardsicherheitspraktiken.

2.1.2 Intrusion Detection System (IDS)

Aufbauend auf der Definition von Intrusion Detection bezeichnen die Autoren ein IDS somit Software, die den Prozess, einen Eingriff zu erkennen, automatisiert [20]. Nach Jones et al. [12] werden Informationssysteme immer umfassender werden und haben einen stetig höher werdenden Wert für Netzwerksicherheit. Schneier [22] bezeichnet IDS vereinfacht als das Netzwerkäquivalent zu Virenschaltern. IDS untersuchen den Netzwerkverkehr oder die auf den Hosts laufenden Prozesse auf Anzeichen für einen Angriff. Wenn sie einen solchen erkennen, schlagen sie Alarm. Seiner Ansicht nach sind IDS dabei nicht darauf ausgelegt, dass jeder Angriff identifiziert wird. Da es immer ein Angriffsszenario gibt, in dem ein ausreichend geschickter Angreifer ein Erkennungssystem umgehen können wird. Trotzdem stellt ein IDS eine wirksame Sicherheitsmaßnahme als Teil der Detektionsphase von NSM dar.

3 Voraussetzungen und verwandte Arbeiten

3.1 Kontextbegriff

Kontext ist zwar ein Begriff, unter dem sich die meisten Menschen intuitiv etwas vorstellen können, ihn zu erläutern oder gar korrekt und vollständig zu definieren, fällt allerdings um ein Vielfaches schwerer [8]. Auch herrschen in verschiedenen Fachbereichen jeder davon mit anderen Sachverhalten und Problemen, welche die jeweiligen Autoren versuchen zu lösen, abweichende Auffassungen darüber, welche Ansprüche eine Definition erfüllen muss. Das erschwert eine allumfassende, konkrete Bestimmung zusätzlich [5]. Ziel der Arbeit ist also nicht, Kontext abschließend zu definieren. Dies ist in Anbetracht der vielen verschiedenen Ansätze und dem fehlenden Konsens in der Literatur darüber, wie solch eine Definition aussehen soll [4, 24] nicht zufriedenstellend möglich. Die Menge an Anwendungsfällen und damit auch die zu beachtenden Gesichtspunkte, die man für eine allen Ansprüchen genügende Definition benötigt, sind dafür zu umfangreich. Stattdessen wird versucht, mithilfe relevanter Darlegungen anderer Autoren die historische Entwicklung des Kontextbegriffs für den Bereich der Informatik im Allgemeinen und der Zugriffskontrolle im Speziellen darzustellen. Dies soll ein Verständnis dafür schaffen, auf welchen Grundlagen, Ansätzen und Ideen der Kontextbegriff dieser Arbeit entstanden und aufgebaut ist. Dies ermöglicht dem Leser eine Einordnung davon, wie Kontext und Kontextsensitivität verwendet werden und er kann einen Abgleich mit seiner eigenen Interpretation der Begrifflichkeiten durchführen.

3.1.1 Historische Entwicklung

Die Definition von Kontext ist auch in der Literatur seit jeher ein Thema. Eine der frühesten von Schilit et al. [21] bestimmt als drei Hauptaspekte, an welchem Ort, in Gegenwart welcher anderen Personen und in der Nähe welcher Ressourcen sich ein Nutzer befindet. Des Weiteren beinhaltet nach [21] Kontext Attribute wie Beleuchtung, Lautstärke, den Grad der Netzwerkverbindung, Kommunikationskosten, und die soziale Situation.

Eine der seit ihrer Entstehung in 2001 am häufigsten zitierten Definitionen stammt von Dey [8]. Laut dieser ist Kontext "jede Information, die genutzt werden kann, um die Situation einer Entität zu charakterisieren. Eine Entität ist eine Person, ein Objekt oder ein Ort mit Relevanz für die Interaktion zwischen Nutzer und Anwendung. Das schließt auch Nutzer und Anwendung selbst mit ein". Sie wird allgemein hin von den vielen anderen Autoren als Quasikonsens akzeptiert [3, 4, 24] oder sogar als Ausgangspunkt für ihre eigene Definition genutzt [13, 27].

Kaltz et al. [25] verstehen Kontext als einen Kontextraum, also eine Kombination aus Kontextparametern, Elementen einer domänenspezifischen Ontologie und Dienstleistungsbeschreibungen in Form von $C = \{U; P; L; T; D; I; S\}$. Dabei ist U das Set aus Nutzern und den dazugehörigen Rollen, P die Prozesse und Aufgaben, L der Ort, T der Zeitfaktor, D beschreibt das Gerät, I die verfügbaren Informationen und S die verfügbaren Dienstleistungen. Ein spezifischer Kontext ist somit ein Punkt in diesem Raum.

Einen anderen Ansatz wählen Bazire und Brézillon. Sie haben 150 Kontextdefinitionen analysiert und sind dabei zu der Erkenntnis gekommen, dass Kontext wie eine Begrenzung funktioniert, welche das Verhalten eines Systems, Nutzers oder Computers in einer bestimmten Tätigkeit beeinflussen. Allerdings herrscht ihrer Ansicht nach kein Konsens darüber, ob Kontext extern oder intern, ein Set aus Informationen oder Abläufen, statisch oder dynamisch ist [5].

Im Bezug auf kontextsensitive Zugriffskontrolle sind die Arbeiten von Kayes et al. [13, 14, 15] erwähnenswert. In [13] definieren diese Kontextinformationen, in Bezug auf Zugriffskontrollentscheidungen, als relevante Informationen über den Zustand einer Entität (Nutzer, Ressource, Ressourcenbesitzer) und deren Umgebung oder die Beziehung zwischen Entitäten.

3.2 Kontextkategorisierung

Genauso relevant für diese Arbeit wie eine Übersicht gängiger Definitionen ist ein Überblick darüber, wie Kontext kategorisiert werden kann. Nach dem Kontextverständnis in der Literatur wird also im Folgenden die Kategorisierung von Kontext näher betrachtet. Dazu erfolgt in diesem Kapitel eine Vorstellung verschiedener ausgewählter Unterteilungsansätze. Auf diese wird sich auch bei der Erstellung der Taxonomie dieser Arbeit in 4 bezogen.

3.2.1 Übersicht relevanter bereits vorgeschlagener Kategorien

Auch bei der Kategorisierung von Kontext gibt es richtungsweisende Vorschläge:

Abowd et al. [2] haben einen der wegweisenden Mechanismen zur Definition von Typen von Kontext vorgeschlagen. Sie identifizierten Ort, Zeit, Identität und Aktivität als primäre Kontexttypen. Weiterhin wird sekundärer Kontext als etwas definiert, das durch Nutzung von Primärkontext erschlossen werden kann.

Schilit et al. [21] kategorisieren Kontext basierend auf drei Fragen, die genutzt werden können, um den Kontext zu bestimmen, in drei Kategorien:

1. Informationen, die sich auf einen Ort beziehen, beispielsweise GPS-Koordinaten, Bezeichnungen von Institutionen oder Gebäuden (ein Café, ein Krankenhaus, eine Universität), spezifische Namen (z. B.: Technische Universität Dresden) spezifischen Adressen (z. B.: APB Nöthnitzer Str. 46) oder Nutzerpräferenzen (z. B. das Lieblingsrestaurant eines Nutzers)
2. Informationen über andere Personen, die in der Nähe aufhalten
3. Informationen darüber, welche Ressourcen (Maschinen, technische Geräte, Betriebsmittel) sich im direkten Umfeld eines Nutzers befinden

Henricksen [11] ordnet Kontext basierend auf der betrieblichen Kategorisierungstechnik in 4 verschiedene Kategorien:

1. Messbar: Informationen, die aus unmittelbar messbaren Werten bestehen. Diese ändern sich oft oder gar kontinuierlich.
2. Statisch: Informationen, die sich während der Lebenszeit eines Systems gleich bleiben.
3. Profiliert: Informationen, die sich selten ändern.
4. Abgeleitet: Informationen, die unter Verwendung anderer Daten gewonnen wurden.

Van Bunningen et al. [23] ordnen Kategorisierungsversuche in zwei übergeordnete Gruppen: Betrieblich und Konzeptionell.

1. betriebliche Kategorisierung: Einordnung anhand dessen, wie der Kontext akquiriert, modelliert und behandelt wird.
2. konzeptionelle Kategorisierung: Einordnung anhand der Bedeutung des Kontextes und der konzeptionellen Beziehungen

Chong et al. [6] schlagen Historie als Kontextkategorie vor. Dabei wird die zeitliche Entwicklung einer bestimmten Messgröße als Kontext definiert. Das erlaubt die Festlegung eines Normalzustandes für dieser Größe. Das ermöglicht laut den Autoren unter Umständen eine Vorhersage darüber, welche Werte die Messgrößen zukünftig annehmen könnten.

4 Design

In diesem Kapitel soll es darum gehen, welche Punkte bei einem Entwurf einer kontextsensitiven Lösung zur Leistungsverbesserung und Problembehebung eines IDS beachtet werden sollte. Dazu wird zuerst die Notwendigkeit einer Kontexttaxonomie erläutert. Dann erfolgt die Definition der einzelnen Kategorien um die Taxonomie dieser Arbeit zu konstruieren. Danach wird festgelegt, in welcher Form die Kontextinformationen in den einzelnen Kategorien vorliegen müssen und wie man diese Informationen aus unterschiedlichen Quellen sammelt um damit die Kategorien zu füllen. Zusätzlich wird noch darauf eingegangen, welche Anforderungen ein IDS erfüllen sollte, welche davon im speziellen für diese Arbeit relevant sind und welche Schwächen des IDS mithilfe des später in diesem Abschnitt vorgeschlagenen Designs und Kontextsensitivität gelöst werden.

4.1 Notwendigkeit der Erstellung einer Taxonomie

Im Abschnitt 3 wurden viele verschiedene Schemata zur Kategorisierung vorgeschlagen, keines dieser Schemata ist aber ausreichend um den Ansprüchen dieser Arbeit gerecht zu werden. Auch eine Evaluation 16 verschiedener Arbeiten durch Perera et al. [18] legt nahe, dass keine Kategorisierung allein allen Anforderungen hinreichend gerecht werden kann. Deshalb wird also statt ein einzelnes Schema zu verwenden, eine Kombination bereits vorgeschlagener Kategorisierungsschemata zum Ausgleich der Nachteile der einzelnen Bestandteile vorgeschlagen und diese mit einer eigenen Kategorisierung aus der Perspektive eines Netzwerkes kombiniert. Dazu werden die Kategorien zuerst im Bezug auf kontextsensitive Zugriffskontrolle definiert. Einige der verwendeten Kategorien bauen dabei auf den im Kapitel 3 vorgestellten Ansätzen auf und sind dementsprechend gleich benannt. Abhängig davon, mit welchem Fokus sie der jeweilige Autor konstruiert hat, wurden sie für diese Arbeit mehr oder weniger stark angepasst.

4.2 Erstellung einer Taxonomie

Die in dieser Arbeit präsentierte Taxonomie besteht aus 3 Komponenten. Einer Unterteilung mit Fokus auf menschliche Nutzer und Menschen generell, eine Kategorisierung aus dem Betrachtungswinkel eines Computernetzwerkes und einer Historie, um die zeitliche Entwicklung der beiden Schemata einzuordnen und damit die Aussagekraft zu erhöhen. Die Abbil-

dung 4.1 und Abbildung 4.2 ¹ veranschaulichen dabei die Beziehung der Unterkategorien zueinander und sind jeweils Beispiele dafür versehen, welche Kontextinformationen in den einzelnen Kategorien vorkommen können.

4.2.1 Mensch

Eine naheliegende Sichtweise zur Kategorisierung von Kontext, welche auch von vielen Autoren, auf die im Kapitel 3 eingegangen wurde, verwendet wird, ist die aus der Perspektive des Nutzers, also einer Person. Für eine Einordnung im Rahmen der Zugriffskontrolle lässt sich ebenfalls argumentieren, dass ein Fokus auf Menschen sinnvoll ist. Zwar rücken gerade in diesem Bereich Personen gelegentlich in den Hintergrund und der überwiegende Teil des Netzwerkverkehrs wird ausgelöst, ohne das Nutzer davon etwas mitbekommen. Aber in letzter Instanz sind sowohl die Verursachenden, die Überwachenden als auch die Forschenden menschlich. Menschen haben vielfältige Anliegen, müssen unauffälligen von auffälligem Netzwerkverkehr unterscheiden oder versuchen die Beweggründe anderer indirekt zu erschließen, zu kategorisieren und in Taxonomien zu visualisieren.

Konzeptionell

Eine Kategorisierung anhand der Bedeutung des Kontextes und der begrifflichen Beziehungen. Bedeutung und Beziehungen werden zusätzlich weiterhin in primäre und sekundäre Kontextinformationen unterteilt.

Primär Primärer Kontext ist jeglicher Kontext, der ohne Verwendung vorhandener Kontextinformationen und ohne Kombination anderer Informationen direkt gewonnen werden können [2].

Sekundär Sekundärer Kontext umfasst alle Kontextinformationen, die mit Hilfe der Verarbeitung von primärem Kontext erschlossen werden können. Realisiert werden kann dies durch die Kombination einzelner Datenpunkte einer oder mehrerer Kategorien oder Abfragen weiterer Daten unter Verwendung der primären Informationen [2].

Betrieblich

Wie schon in der Analyse bereits vorgeschlagener Kategorien, erläutert in Abschnitt 3.2.1, meint betriebliche Kategorisierung eine Einordnung anhand dessen, wie der Kontext akquiriert, modelliert und behandelt wird [23].

Dabei wird weiter in Zeit, Ort, Identität, Aktivität und Grund unterschieden. Die Zeit gibt an wann eine Zugriffsanfrage erfolgt, der Ort von welcher logischen Position von außerhalb oder innerhalb eines Netzwerkes oder von welcher physischen Position in der realen Welt eine Zugriffsanfrage oder Netzwerkverkehr generell stammt, die Identität ist der Indikator dafür welche Entität Zugriff auf eine Ressource im Netzwerk oder Zugang zum Netzwerk erfragt, die Aktivität beinhaltet das was in einer Situation geschieht bzw. welche Aktion eine Entität ausführt.

Die Kategorie Grund enthält alle Informationen darüber warum etwas von einer Entität getan wird oder wurde. In Abbildung 4.1 ist kein Beispiel für einen primären Grund vorhanden, da dieser sowie er in dieser Arbeit auf Basis der Erkenntnisse von Abowd et al. [2] verwendet

¹Als Inspiration für den Stil und die Struktur der Grafiken diente eine in [18] verwendete Abbildung. Diese wurde inhaltlich adaptiert, um für den Kontext der Netzwerksicherheit verwendet werden zu können.

wird aus Zeit, Ort, Identität und Aktivität abgeleitet und deshalb nicht direkt im Netzwerkverkehr oder an anderer Stelle vorliegt. Im Gegensatz zu Ort, Zeit, Identität oder Aktivität ist der Grund dafür, warum eine Aktion ausgeführt wird weder objektiv quantifizierbar noch unterliegt er einem zeitlichen Zusammenhang oder folgt einer zeitlichen Regelmäßigkeit. Deshalb ist es weder möglich noch sinnvoll, den passiven Teil der Historie, also Abtastrate oder Änderungsrate anzugeben und wurde deswegen in den Beispielen in der Abbildung 4.1 unterlassen.

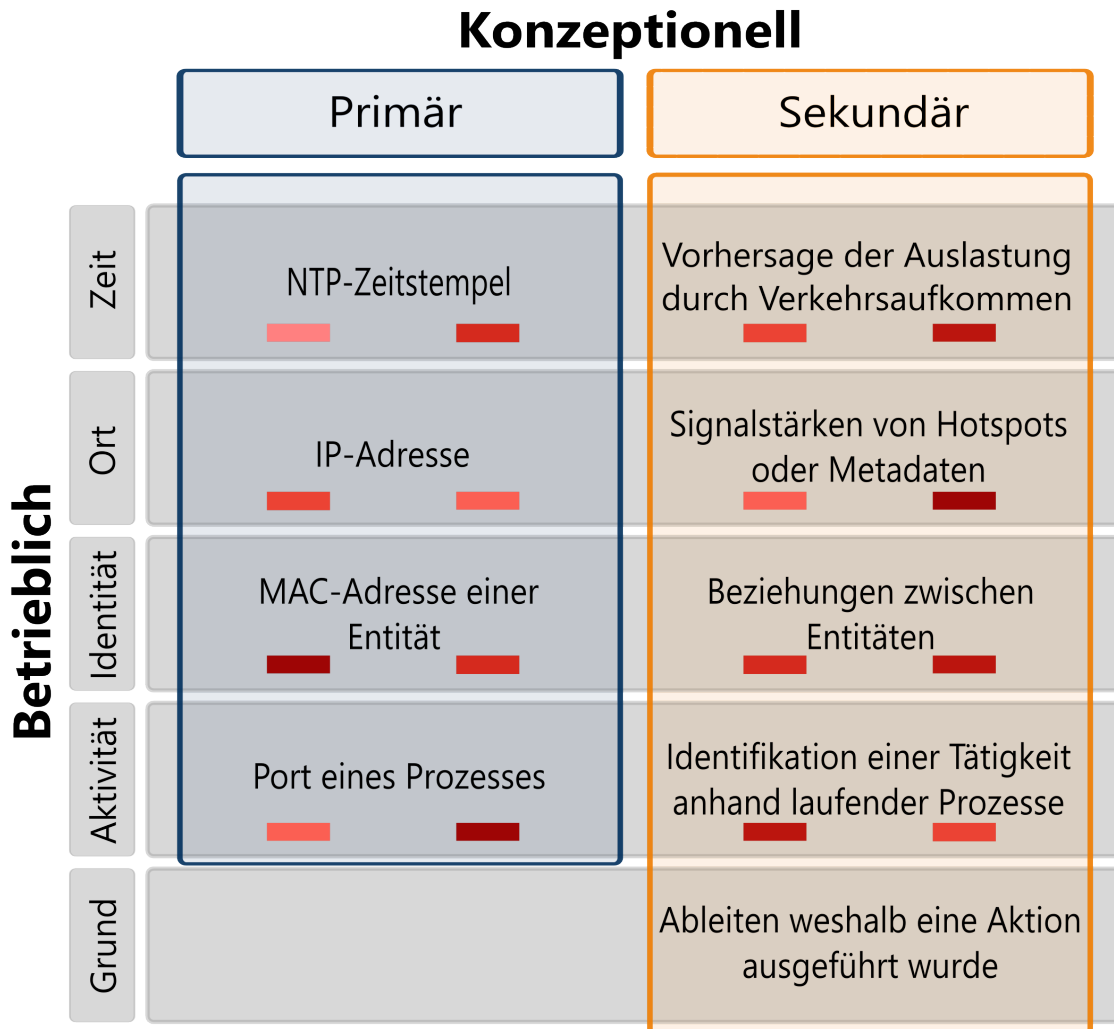


Abbildung 4.1: Nutzerzentrierte Kategorisierung

4.2.2 Netzwerk

Ein Netzwerk besteht im Allgemeinen aus Entitäten, die daran teilnehmen, wie etwa Routern, Switches, Servern oder Endgeräten, Kommunikation zwischen den Entitäten, beispielsweise Anfragen von Endgeräten an Server oder Weiterleitung von Daten zwischen verschiedenen Routern und Annahmen bzw. Normen bezüglich des Verhaltens und der Eigenschaften der Entitäten sowie der Form und dem Inhalt der Kommunikation, so gibt es allgemein Obergrenzen für Verbindungsanfragen oder Mindestanforderungen für Softwareversionen und eine DNS-Anfrage oder das Herunterladen von Dateien muss auf eine durch das dafür verwendete Protokoll festgelegte Art und Weise mit dafür verpflichtend benötigten Informationen geschehen, um erfolgreich zu sein. Im Rahmen der Zugriffskontrolle wird dabei eine

Grundmenge an Normen in Form von Policies initial festgelegt und im Verlauf der Lebenszeit des Netzwerkes mithilfe neuer Erkenntnisse angepasst.

Protokolle

Die Kategorisierung, die anhand des verwendeten Kommunikationsprotokolls erfolgt, orientiert sich am ISO/OSI-Referenzmodell [7]. Die Zuordnung zu einer bestimmten Ebene und damit Kategorie wird dabei durch die für die einzelnen Schichten üblichen Protokolle umgesetzt. Das ermöglicht die Identifikation von Entitäten durch von ihnen genutzte Protokolle. So kann beispielsweise ein Switch oder Router, welcher nur in geringem Maße, in bestimmten Fällen zum Beispiel durch ein Webinterface, Verkehr oberhalb der Vermittlungsschicht generiert, von einem Nutzer oder einer Anwendung, die vermehrt Verkehr erzeugen der Protokolle die für die Anwendungsschicht üblich sind verwendet, unterschieden werden.

Die Anwendungskategorie beinhaltet alle Datenpakete des Netzwerkverkehrs, die zum Beispiel DNS, HTTPS, DHCP oder LDAP als Protokolle einsetzen und somit der Sitzungsschicht, Darstellungsschicht oder Anwendungsschicht zugerechnet werden können.

Die Transportkategorie umfasst Pakete, die Protokolle wie TCP oder UDP verwenden.

Die Vermittlungskategorie enthält Netzwerkverkehr, der Protokolle wie IP oder ICMP zur Kommunikation benutzt.

Entitäten

Einordnung von Kontextinformationen abhängig davon, welche Art von Entität oder Entitäten sie betreffen. Diese Unterscheidung setzt genauso wie die Historie eindeutig identifizierbare Entitäten voraus.

Gerät Informationen, die sich auf ein Gerät beziehen. Geräte sind dabei zum Beispiel der PC eines Nutzers oder ein Router im Netzwerk mit den jeweiligen für sie spezifischen Hardwarekomponenten und deren Firmware bzw. Software. Informationen umfassen beispielsweise die Liste an installierten Anwendungen, die Menge laufender Prozesse oder die Auslastung des Systems oder einzelner Teile.

Nutzer/Rolle Informationen, die sich auf einen Nutzer und dessen Aktionen oder seine Rolle beziehen. Ein Nutzer ist dabei eine physische Person und kann mehrere verschiedene Rollen innehaben. Informationen können beispielsweise Zugriffsrechte, die mit einer Rolle eines Nutzers verbunden sind oder seine Browserhistorie sein.

Anwendung Informationen, die sich auf eine einzelne Anwendung oder eine Kombination aus mehreren Programmen oder Prozessen beziehen, die für sich oder im Zusammenspiel miteinander einen eindeutigen Zweck erfüllen. Dabei wird charakteristischer Netzwerkverkehr erzeugt, also bestimmte Daten nach spezifischen Mustern versendet oder empfangen.

Policies

Einordnung der Kontextinformationen abhängig davon, ob sie den für den Anwendungsfall definierten Normen entsprechen. Dabei gibt es Kommunikation und Verhalten, welches so erwartet wird, also den im Netzwerk und geltenden Konventionen und aufgestellten Regeln folgt und ungewöhnliches, welches in Kombination mit den vorherrschenden Gegebenheiten entweder auffällig oder irrational ist.

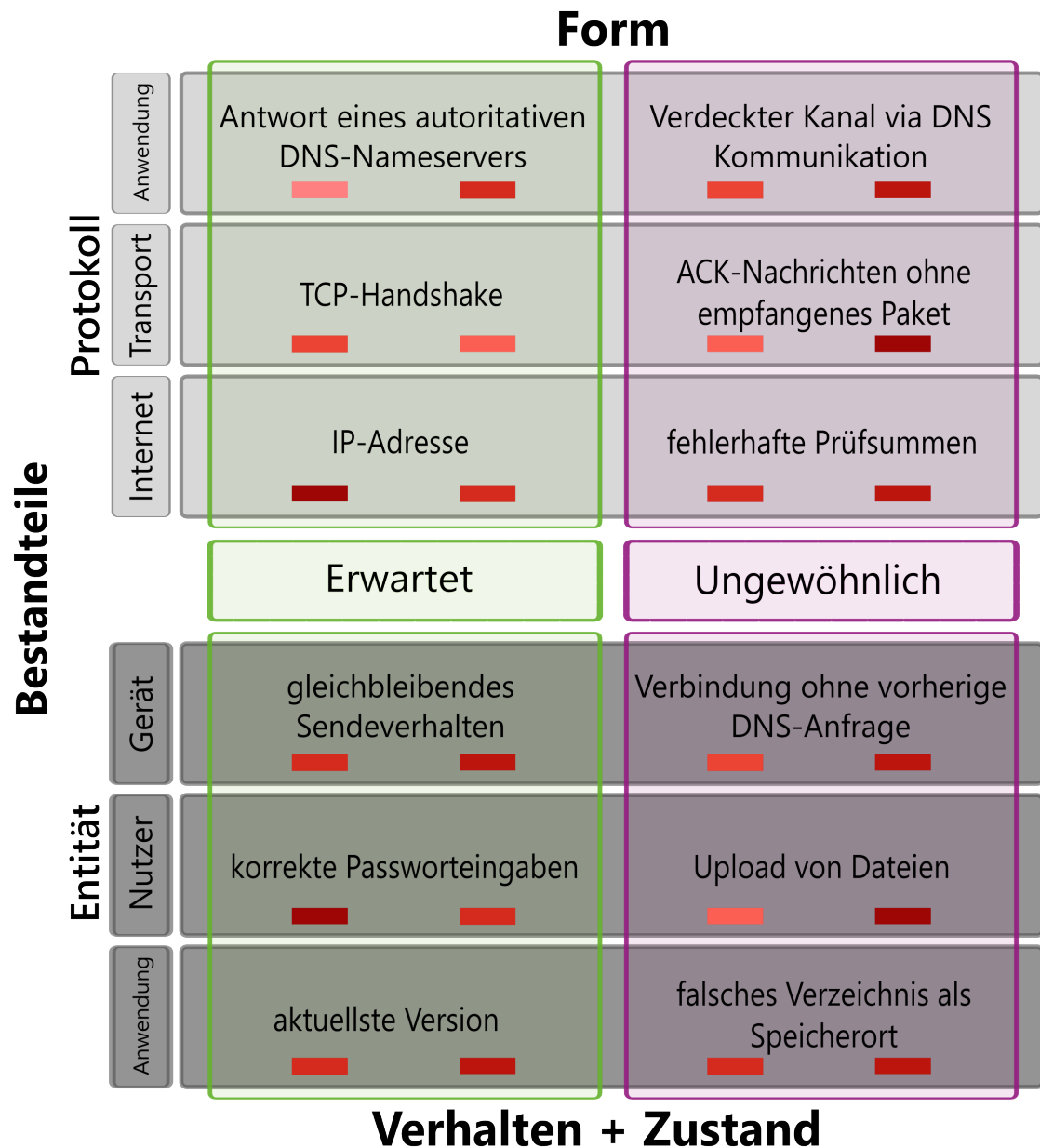


Abbildung 4.2: Kategorisierung von Kontextinformationen als Bestandteile eines Netzwerkes

4.2.3 Historie der Werte einer Kategorie

Unabhängig vom gewählten Fokus oder Blickwinkel auf die Kategorisierung von Kontext benötigt man, um bereits gesammelte Kontextinformationen nutzen zu können, eine Art Gedächtnis. Im Fall eines IDS, welches ohnehin aufgrund seiner Funktionsweise über einen längeren Zeitraum Informationen verarbeitet und neue Erkenntnisse speichert, bietet sich eine Historie, ähnlich wie in Kapitel 3 beschrieben, an. Die Historie einer Entität wird in dieser Arbeit zweigeteilt. Sie besteht aus einem aktiven und einem passiven Teil, also ihrem Verhalten, beispielsweise früheren Verbindungen oder Verbindungsanfragen und dem Zustand der für sie charakteristischen Attribute wie etwa einem Nutzernamen oder die Versionsnummer eines bestimmten Programms. Durch welche Informationen Verhalten repräsentiert wird, ist im Absatz 4.3.1 ausführlich erläutert, die Zusammensetzung des Zustandsteils wird durch die Kenngrößen und deren Änderungsverhalten dargestellt und im Abschnitt Dynamik und

im Abschnitt Kenngrößen genauer spezifiziert. In der Abbildung 4.3 wird veranschaulicht wie dieser Teil in den Abbildungen 4.1 und 4.2 zu verstehen ist.² Insofern es sich um die Historie einer Entität oder eine Entität betreffende Eigenschaften handelt, wird die implizite Annahme getroffen, dass diese Entitäten eindeutig identifizierbar sind, da sonst eine Zuordnung nicht möglich ist.

Dynamik

Es liegt in der Natur der Sache, dass sich die Messwerte, aus denen sich die Historie einer Entität zusammensetzt, je nachdem welchem Teil sie zugeordnet werden, verschieden oft ändern. Statische Messgrößen ändern dabei ihre Werte nie oder nur sehr selten. Dynamische Messgrößen hingegen sehr oft. In diesem Fall wird eine Unterteilung in jährlich, monatlich, wöchentlich, täglich, stündlich, minütlich und sekundlich vorgenommen.

Kenngrößen der aktiven Historie

Die Historie einer Kontextinformation ist weiterhin in zwei verschiedene Kenngrößen, die Änderungsrate und die Abtastrate unterteilt. Die Änderungsrate gibt die Häufigkeit an, mit der eine Wertänderung im Regelfall vorkommt, die Abtastrate wie oft Kontextinformationen vom IDS abgefragt und ggf. aktualisiert werden.

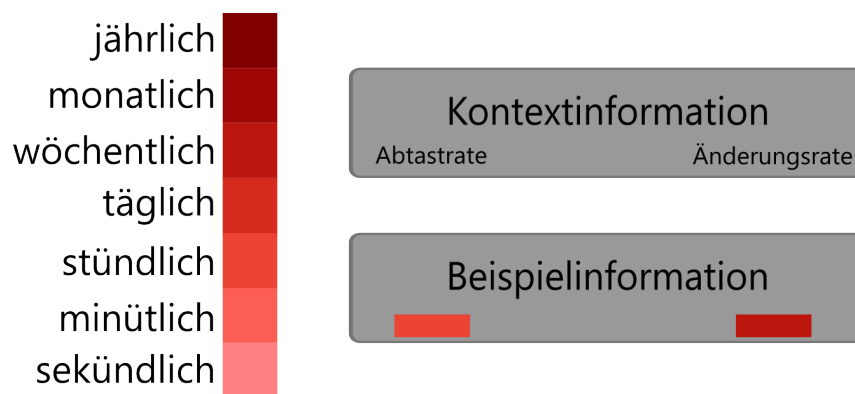


Abbildung 4.3: Dynamik einer Kontextinformation

4.3 Form des Kontextes

Nachdem die einzelnen Kategorien definiert wurden, muss festgelegt werden in welcher Form die Kontextinformationen vorliegen sollen bzw. gebracht werden müssen. Für fast alle Kategorien ist das selbsterklärend. Protokollkontext hat eine durch das Protokoll selbst vorgegebene feste Form, ein Zeitstempel ebenso. Größtenteils ist anhand der Spezifikation der jeweiligen eingeordneten Informationen offensichtlich welche Form die Information haben sollte und die Form des Kontextes lediglich nebensächlich. Deshalb wird nicht auf alle Kategorien separat eingegangen.

Die Historie bildet eine Ausnahme. Hier gibt es je nach Bedarf des Anwendungsfalls einen gewissen Spielraum. Die gespeicherten Informationen werden abhängig von der Implementierung eventuell nicht vom IDS eingelesen oder gar selbst vom IDS erzeugt. Auch gibt es

²Es handelt sich bei den Informationen in den Abbildungen lediglich um Beispiele zur Veranschaulichung des Aufbaus. Die Kombinationen aus den Werten und deren aktiven Teil weicht in der Praxis ab.

höchst unterschiedliche Ansprüche an Form, Umfang und Zeitraum oder Beschränkungen hinsichtlich Performanz und Speicherbedarf.

4.3.1 Historie

Die Historie muss genug Informationen enthalten, um damit neue Urteile in der Gegenwart oder Zukunft zu fällen und Entitäten eindeutig identifizieren zu können. Schließlich kann eine Historie, die dabei helfen soll, Entscheidungen zu treffen, die den Netzwerkverkehr von Entitäten betreffen kann nicht ohne eindeutig identifizierte Entitäten funktionieren.

Die Identifikation erfolgt anhand der Headerinformationen der einzelnen Kommunikationsschichten. In der Historie gespeichert werden entweder die gesamte Payload eines Pakets oder zumindest ein ausreichend aussagekräftige Teilmenge. Die Regelmäßigkeit mit der die Historie aktualisiert bzw. erweitert wird ist von den jeweiligen Gegebenheiten und Anforderungen abhängig. Kommunikationspartner benötigen eine Grundmenge an Informationen um miteinander kommunizieren zu können. Genauso benötigt ein IDS eine bestimmte Mindestmenge an Informationen, um initial eine Entscheidung treffen zu können, zu finden sind die dafür benötigten Informationen in Tabelle 4.1.

Tabelle 4.1: Benötigte Informationen der Historie einer Entität

Identifikation	Rekonstruktion der ursprünglichen Zugriffsentscheidungen
MAC-Adresse	MAC-Adresse des Ziels
IP-Adresse	IP-Adresse des Ziels, gesetzte Flags, Protokoll, Gültigkeitsdauer
Port	Port des Ziels, protokollspezifische Informationen

Wenn nicht wenigstens diese Information im Netzwerkverkehr enthalten sind, kann ein IDS eine Entität weder eindeutig zuordnen noch eine fundierte Entscheidung treffen. Da also mindestens diese Datenlage bereits benötigt wird, erscheint sie auch als sinnvolle Mindestanforderung für eine Historie. Ohne diese Daten kann keine ausreichende Rekonstruktion des Kenntnisstandes, mit dem die Zugriffsentscheidungen ursprünglich getroffen wurden, ermöglicht werden. Damit ist also keine Aussage über vergangenes Verhalten einer Entität und dessen Bewertung möglich.

4.4 Kontextgewinnung

Jegliche Kategorisierung von Kontext ist von geringem Nutzen, ohne das Informationen vorhanden sind, die kategorisiert werden können. Dabei bedacht werden sollte, an welcher Stelle und wie oft Informationen abgerufen oder automatisch aktualisiert werden müssen [18]. Pimenidis et al. [19] nennen KMDBs, Scanner oder Crawler als Möglichkeiten Kontextinformationen zu sammeln. Zusätzlich dazu sind Anwendungen die direkt auf Endgeräten laufen und von dort aus Daten an das IDS schicken erwähnenswert. Eine Übersicht und genauere Erklärung der einzelnen Komponenten findet sich in Tabelle 4.2.

Tabelle 4.2: Akquirierungspunkte für verschiedene Kontextinformationen

Werkzeug	Erklärung
Konfigurationsmanagement-datenbank (KMDB)	Beinhaltet die Konfigurationsparameter aller bekannten Hosts, einschließlich der vollständigen Historie, mit Details zu Hardware, Software, Prozessen, Infrastrukturen, Verantwortlichkeiten und anderen Komponenten.
Scanner für Netzwerkschwachstellen	Scannen von beispielsweise Ports um sich einen Überblick über laufende Anwendungen und mögliche Schwachstellen zu verschaffen.
Crawler	Erlaubt es die auf einem einzelnen Port aktiven Anwendungen zu ermitteln.
Endpunktagent	Ermöglicht das Abfragen verschiedenster Systemparameter direkt vom Hostsystem. Die Menge an abfragbaren Informationen hängt dabei stark von der gewählten Implementierung ab.

Je nach Wichtigkeit des jeweiligen Hosts variiert der Detailgrad der Informationen die in einer KMDB vermerkt sind. Bei kritischen Infrastrukturen wie Routern, Switches und zentralen Servern sollte die KMDB mindestens das Betriebssystem, dessen Version sowie alle Anwendungen mit ihren Versionen und Patch-Ständen enthalten. Bei allen hier aufgelisteten Werkzeugen sollte bedacht werden, dass aufgrund der Fehleranfälligkeit dieser Herangehensweise Daten zu sammeln, die gewonnenen Informationen in jedem Fall von Menschen gegen geprüft werden sollten.

4.5 Umwandlung der Taxonomie in IDS-Signaturen

Um die im Abschnitt 4.2 festgelegten Kategorien in für ein IDS nutzbare Form zu bringen, gilt es gewisse Dinge zu beachten:

Die Kontextsensitivität eines Computers unterscheidet sich drastisch von der eines Menschen. Rechensysteme sind sehr gut darin, Daten zu erfassen und zu sammeln, aber Menschen sind immer noch nötig, um verschiedene Kontexte zu erkennen und zu entscheiden, welches Handeln in einer bestimmten Situation angemessen ist [8]. Der limitierende Faktor des Potenzials eines kontextsensitiven Systems ist das Maß an Kontext, das ein Entwickler vorhersehen und codieren kann. Es ist aber weder beim Design noch später bei der Implementierung schwer möglich, alle Zusammenhänge vorherzusehen. Dementsprechend schwer ist es, ein in sich geschlossenes und allumfassendes Regel-Set festzulegen [18]. Nach Greenberg et al. [10] gibt es drei nicht-triviale Hauptaspekte, die beim Entwerfen eines kontextsensitiven Systems beachtet werden sollten:

1. Spezifizieren aller möglichen Kontextzustände.
2. Wissen welche Informationen einen konkreten Kontextzustand akkurat festlegen.
3. Welche Aktion im jeweiligen Zustand ausgeführt werden soll.

Diese Fragen sollten sich im Verlauf des Entwurfes eines jeden Designs gestellt werden und waren auch für diese Arbeit von hoher Relevanz, da sie eine große Menge an aufkommenden Problemen greifbar und damit leichter lösbar machen.

4.5.1 Entscheidung für ein IDS

Milenkoski et al. [17] kategorisieren IDS, wie in Tabelle 4.3 zu sehen, anhand der überwachten Plattform der verwendeten Erkennungsmethode und der Struktur, in der sie eingesetzt werden.

Tabelle 4.3: Einteilungsansätze für IDS

Eigenschaft	IDS Typ	Beschreibung
Plattform	Host	Überwacht Aktivitäten auf dem System, auf dem es eingesetzt wird, um lokale Angriffe zu erkennen.
	Netzwerk	Überwacht Aktivitäten im Netzwerk um Angriffe, die über eine Netzwerkverbindung ausgeführt werden, zu erkennen.
	Hybrid	kombiniert host- und netzwerk-basierte Intrusion Detection Systeme.
Angriffserkennung	Signatur	Vergleicht System- und/oder Netzwerkaktivitäten mit einem Set von Signaturen bekannter Angriffe. Daher ist es nicht in der Lage, Zero-Day-Angriffe zu erkennen, d. h. Angriffe, die Schwachstellen ausnutzen, die vor der Ausführung der Angriffe nicht öffentlich bekannt sind.
	Anomalie	Verwendet ein Basisprofil regulärer Netz- und/oder Systemaktivitäten als Referenz, um zwischen regulären und auffälligen Aktivitäten zu unterscheiden, wobei Letztere als Angriffe behandelt werden. Wird typischerweise durch die Überwachung regulärer Aktivitäten trainiert, um ein Basisaktivitätsprofil zu erstellen.
	Hybrid	Verwendet sowohl signatur-basierte als auch anomalie-basierte Angriffserkennungsmethoden.
Struktur	Zentral	Kann nur an einem einzigen Standort eingesetzt werden.
	Verteilt	Besteht aus mehreren Teilsystemen zur Erkennung von Angriffen, die an verschiedenen Standorten eingesetzt werden können und miteinander kommunizieren, um für die Erkennung von Angriffen relevante Daten, wie z. B. Angriffswarnungen auszutauschen und so koordinierte Angriffe auf mehrere Standorte in einer bestimmten zeitlichen Abfolge erkennen.

Schwächen der verfügbaren IDS

Der Hauptnachteil eines signaturbasierten IDS ist ausschließlich Angriffe zu erkennen, die vordefinierten Verhaltensmustern entsprechen. Der Hauptnachteil eines anomaliebasierten IDS ist die Baseline. Zuerst muss eine Baseline festgelegt werden, was mitunter je nach Größe und Struktur des Netzwerkes sehr komplex und damit fehleranfällig sein kann. Zusätzlich muss diese Baseline mitunter sehr oft aktualisiert werden, um auf Änderungen im Netzwerk zu reagieren. Dies verlangt dem Überwachenden zusätzliche Ressourcen ab.

4.6 Anspruch an das IDS

Um die Performance des konkret gewählten IDS verbessern zu können, muss festgelegt werden, welche Kriterien die Leistung beeinflussen bzw. bestimmen und wie man diese gewichtet. Mell et al. [16] geben eine Übersicht über verschiedene Charakteristiken zur quantitativen Bestimmung der Erkennungsgenauigkeit eines IDS und erläutern zusätzlich die Wechselwirkungen zwischen einzelnen Kriterien, die beim Vergleich verschiedener IDS-Lösungen beachtet werden sollten.

Abdeckung

Diese Metrik gibt an, welche Typen von Angriffen ein IDS unter idealen Bedingungen erkennen kann.

Wahrscheinlichkeit falscher Alarme

Gibt die Wahrscheinlichkeit, dass durch ein IDS ausgelöste Alarme durch gutartigen bzw. nicht-schädlichen Netzwerkverkehr verursacht wurden, an.

$$\text{Rate an falsch - Positiven Meldungen} = \frac{\text{Anzahl falscher Alarme}}{\text{Anzahl aller Alarme}}$$

Wahrscheinlichkeit einer Erkennung

Gibt die Rate der durch das IDS korrekt erkannten Angriffe an.

$$\text{Erkennungswahrscheinlichkeit} = \frac{\text{Anzahl korrekt erkannter Angriffe}}{\text{Anzahl aller Angriffe}}$$

Resistenz

Ein IDS und der menschliche Administrator hinter dem System weisen Probleme auf, die nicht direkt beim Umgang mit verarbeitetem Netzwerkverkehr, sondern schon bei der bewussten, unbewussten oder erzwungenen Entscheidung, welcher Netzwerkverkehr überhaupt infrage kommt, entstehen:

1. Eine zu große Menge an zu verarbeitendem Netzwerkverkehr, die die Verarbeitungskapazität eines IDS übersteigt, kann dazu führen, dass Netzwerkpakete verworfen und Angriffe nicht erkannt werden.
2. Pakete die zwar nicht böartig sind, aber so konstruiert sind, dass sie möglichst viele IDS Alarme auslösen, überfordern den Administrator oder stören eventuell sogar die Verarbeitung von Paketen generell.
3. Ein Angreifer könnte eine Vielzahl "harmloserer", aber trotzdem noch als schädlich zu deklarierende Pakete senden, um einen größeren Angriff im Netzwerkverkehr zu verschleiern.
4. Pakete, die möglicherweise vorhandene Fehler im IDS selbst ausnutzen.

Korrelation zwischen Einzelereignissen

Korrelation demonstriert wie gut ein IDS einen Zusammenhang zwischen einzelnen Events, möglicherweise verschiedenen Ursprungs, herstellen kann. Die Ereignisse können dabei aus Routern, Firewalls, Anwendungen, dem IDS selbst oder einer großen Bandbreite anderer Quellen stammen.

Vorhersehen von unbekannten Angriffen

Gibt an, wie gut ein IDS einen Angriff erkennt, der so noch nicht aufgetreten ist. Signatur-basierte IDS sind allgemein mit wenigen Ausnahmen nicht in der Lage, solch einen Angriff zu erkennen. Normalerweise erhöht die Fähigkeit eines Systems, einen noch unbekannten Angriff zu erkennen, im Vergleich zu Systemen, die dies nicht versuchen, zusätzlich die Rate an falsch-positiven Meldungen.

Identifizieren von Angriffen

Die Identifikationsfähigkeiten eines IDS geben an, wie gut ein IDS einem Angriff, den es erkennt, einen Namen oder eine Kategorie, beispielsweise ein CVE-Nummer, zuordnen kann.

Beurteilung von Angriffen

Die Beurteilung ist ein Indikator dafür, ob ein IDS den Erfolg und die Auswirkungen eines Angriffes korrekt beurteilen kann. In aktuellen Netzwerkumgebungen schlagen viele Angriffe bzw. Angriffsversuche fehl. Die meisten IDS unterscheiden allerdings nicht zwischen erfolgreichen und fehlgeschlagenen Angriffen. Für denselben Angriff können manche IDS die Anzeichen dafür, ob ein Angriff erfolgreich war, erkennen andere lediglich, dass ein Angriff stattgefunden hat, allerdings ohne feststellen zu können, ob er erfolgreich war. Die Fähigkeit, den Erfolg eines Angriffs festzustellen, ist für die Analyse der Angriffskorrelation und des Angriffsszenarios von wesentlicher Bedeutung; sie vereinfacht auch die Arbeit eines Analysten erheblich, indem sie zwischen wichtigeren erfolgreichen Angriffen und den in der Regel weniger schädlichen fehlgeschlagenen Angriffen unterscheidet. Die Messung dieser Fähigkeit erfordert Informationen über fehlgeschlagene und erfolgreiche Angriffe.

Einordnung der Kriterien

Die Wahrscheinlichkeit, einen Angriff zu erkennen, variiert mit der Rate an falsch-positiven Meldungen. Die Verbesserung der Rate von falsch-positiven Meldungen und Erkennungsrate stehen sich diametral gegenüber. Deutlich wird dies am jeweiligen Extremfall. Erkennt das System den gesamten Netzwerkverkehr als schädlich so erkennt es auch alle Angriffe korrekt. Die Erkennungsrate ist also optimal. Die Falsch-positiv-Rate hingegen sehr schlecht. Erkennt das System kein Paket als Angriff, gibt es keine falsch-positiven Meldungen. Allerdings ist damit auch die Erkennungsrate maximal schlecht. Bei der Konfiguration eines IDS kann also nur in Hinsicht auf eine der beiden Metriken optimiert werden.

4.6.1 Auswahl der Kriterien

Diese Arbeit beschäftigt sich ausschließlich mit sicherheitsrelevanten Leistungsmetriken. Dabei soll hauptsächlich der Einfluss von Kontextsensitivität auf die Abdeckung, die Erkennungswahrscheinlichkeit und die Rate falsch-positiver Alarme betrachtet werden. Ergänzend wird beleuchtet, ob zusätzlicher Kontext die Interpretierbarkeit von Meldungen durch einen menschlichen Nutzer und das Identifizieren von Angriffen verbessert.

5 Implementierung

Der im Kapitel 4 dargelegte Aufbau eines Netzwerkes findet sich auch in der Implementation wieder. Scapy, Wireshark und Zeek verarbeiten Kommunikationsdaten, Zeek-agent stellt die Attribute der Entitäten bereit und mit den in den im Nachfolgenden vorgestellten Skripten werden beispielhaft die geltenden Normen und das dadurch abgedeckte Verhalten festgelegt.

5.1 Erläuterung der wichtigsten Komponenten

Eine kurze Vorstellung der wichtigsten Komponenten der Implementierung, insofern sie im Rahmen der einzelnen Schritte des Versuchsaufbaus essenziell sind.

5.1.1 Zeek

Ein passives, quelloffenes Analysewerkzeug für Netzwerkverkehr, das durch eine eigene Skriptsprache unter anderem die Implementierung beliebiger Analyseaufgaben, die Interpretation von Netzwerkverkehr und Erstellung von Protokollen auf der Grundlage dieses Verkehrs ermöglicht.

Zeek ist dabei weder rein signatur-basiert wie Suricata noch ausschließlich als Protokollanalytiker im Sinne von Wireshark zu verstehen. Vielmehr handelt es sich bei Zeek um eine Mischung, die eine kompakte, aber dennoch detailgetreue Darstellung von Netzwerkprotokollen ermöglicht. Dies erlaubt ein besseres Verständnis des Netzwerkverkehrs und der Netzwerknutzung [1].

5.1.2 Zeek-Agent

Zeek-Agent ist ein Endpunkt-Agent, der Informationen für zentrales Monitoring an Zeek sendet. Abfragen kann man verschiedene Aktivitäten eines Hostsystems, darunter zum Beispiel aktuell laufende Prozesse, offene Sockets oder den Inhalt bestimmter Dateien. Diese erscheinen in Zeek, genauso wie Netzwerkaktivität als Ereignisse und können so in Skripten verwendet werden [26].

5.2 Versuchsaufbau

Der Ablauf für alle Anwendungsfälle ist grundsätzlich sehr ähnlich:

1. Erzeugung
2. Mitschnitt via WireShark
3. Analyse mittels Zeek und Zeek-Agent
 - a) Einlesen von Netzwerkverkehr
 - b) Einbindung des zusätzlichen Kontextes
 - c) Logging

5.2.1 Erzeugung

Der Netzwerkverkehr wurde mit Hilfe von Scapy generiert. Das ermöglicht den für die verschiedenen Szenarien benötigten Netzwerkverkehr zu erzeugen und die einzelnen Schichten eines Pakets an den jeweiligen Anwendungsfall anzupassen. In Abbildung 5.1 ist dieser Prozess ausschnittsweise dargestellt. Eine Übersicht über alle dafür verwendeten Skripte findet sich im Anhang.

Quelltext 5.1: Konfiguration und Versendung eines Pakets

```

6 def send_packet(ip_address_src, ip_address_dst):
7     source_server = ip_address_src
8     target_server = ip_address_dst
9     layer_2 = Ether()
10    layer_3 = IP(src=source_server, dst=target_server)
11    layer_4 = TCP(sport=80, dport=43468)
12    tcp_pkt = layer_2 / layer_3 / layer_4
13    sendp(tcp_pkt)

```

5.2.2 Logging

In Zeek erfolgt das Erstellen eines Logs immer nach demselben Prinzip:

1. Vor dem Ausführen eines Skriptes wird ein Log und die darin zu speichernden Informationen festgelegt (Z. 3-10)
2. Nutzer-definierter Log wird initialisiert (Z. 19)
3. Form des Eintrags für den Log wird definiert (Z. 20)
4. Log-Eintrag wird der Verbindung als Information hinzugefügt (Z. 23)
5. Eintrag wird in Log geschrieben (Z. 24)

Quelltext 5.2: Generierung einer Log-Datei mit Verbindungsinformationen

```

1 export {
2   # Create an ID for our new stream. By convention, this is called "LOG".
3   redef enum Log::ID += { LOG };

5   # Define the record type that will contain the data to log.
6   type Info: record {
7     timestamp: time &log;
8     id: connection_id &log;
9     notice: string &log;
10  };
11 }

13 redef record connection += {
14   # By convention, the name of this new field is the lowercase name
15   # of the module.
16   examplelog: Info &optional;
17 };
18 event zeek_init(){
19   Log::create_stream(ExampleModule::LOG, $columns=Info, $path="examplemodule");
20   local record: ExampleLog::Info = $ts=current_time(), $id=c$id,
21     $notice="Example Notice"];
22   # Store a copy of the data in the connection record so other
23   # event handlers can access it.
24   c$examplelog = record;
25   Log::write(ExampleModule::LOG, rec);
26 }

```

5.3 Skripte

Der Kern der Implementierung sind Zeek Skripte. Hier werden die gesammelten Kontextinformationen verwendet. Nachfolgend werden für einige ausgewählte Kategorien Anwendungsfälle vorgestellt.

5.3.1 Geografische Koordinaten und Ortszeit

Das Volumen von durch Menschen verursachten Netzwerkverkehr ist in der Regel größtenteils tageszeitabhängig. So wird üblicherweise nachts weniger kommuniziert als am Tag. Deshalb erscheint es sinnvoll, den Grenzwert für erzeugtes Verkehrsaufkommen, ab dem ein Sender als potenziell böswillig eingestuft wird, je nach Uhrzeit anzupassen.

Das Skript ordnet eine IP-Adresse zu einem geografischen Ort zu (Z. 32).

Berechnet die Ortszeit am Ursprung der Anfrage mithilfe der lokalen Uhrzeit in Kombination mit dem aus dem Abstand ermittelten Zeitunterschied (Z. 36 - 42).

Passt des Grenzwertes entsprechend an (Z. 45).

Quelltext 5.3: Geolokalisierung und Setzen des Grenzwertes

```

31 function geolocation(c: connection):double{
32     local origin_longitude = lookup_location(c$Id$orig_h)$longitude;
33     return origin_longitude;
34 }

36 function time_at_geolocation(longitude: double): int{
37     local time_to_add = (longitude - home_longitude)*240;
42     return useable_time_at_origin;
43 }

45 function set_threshold(c_time: int): double{
46     if(c_time< opening_time || c_time > closing_time )
47         threshold = threshold-night_time_decrease;
48     else
49         threshold = threshold+day_time_increase;
50     return threshold;
51 }

```

5.3.2 Verwendete Ports

Wenn eine Verbindung oder ein Verbindungsversuch mit einem bestimmten Port des Systems als Ziel beobachtet wird, ohne das es einen Prozess gibt, der diesem Port durch Zeek-Agent zugeordnet werden kann, wird die Verbindung oder der Verbindungsversuch im dazugehörigen Log vermerkt.

Das Skript erfragt bei den Instanzen des Zeek-Agents in regelmäßigen Abständen die auf Hostsystemen laufenden Prozesse und deren verwendete Ports (Z. 77 und 63).

Das Skript vergleicht den Zielport jeder eingehenden Verbindung mit der Liste von verwendeten Ports auf dem Hostsystem (Z. 54).

Abhängig vom Ergebnis wird eine Meldung in den Log geschrieben (Z. 55).

Quelltext 5.4: Abfrage und Abgleich der Ports

```

52 function check_outgoing_connection(c:connection){
53     local _port = port_to_count(c$Id$orig_p);
54     if(_port !in local_ports){{
55         local rec: Querytest::Info = [$ts=current_time(), $id=c$Id, $notice="No
           Application running on this port"];

63 event users_result(ctx: ZeekAgent::Context, data: Columns) {
64     local new_entry : count;
65     local connection_port = data$remote_port;
66     new_entry = connection_port;
67     local_ports[new_entry] = data$name;
68 }

77     local port_query = ZeekAgent::query([$sql_stmt=str_stmt_join,
           $event=query_event, $schedule=_schedule]);

```

5.3.3 DNS-Auflösung

Menschliche Nutzer verwenden bei der Nutzung ihres Endgerätes im Gegensatz zu Computern keine IP-Adressen, um Suchanfragen zu formulieren. DNS ordnet IP-Adressen menschenfreundliche Domainnamen zu. Wenn das System eines Nutzers eine Verbindung aufbaut, geht dem eine DNS-Anfrage von diesem Gerät an einen DNS-Nameserver voraus oder die IP-Adresse ist in lokalen Konfigurationsdateien auffindbar. Wenn beispielsweise eine TCP-Verbindung eines Webbrowsers beobachtet wird, ohne dass eine zuordenbare DNS-Anfrage erfolgt ist oder die IP-Adresse in der Routingdatei des Endgerätes vermerkt ist, ist das in den meisten Fällen ein Grund zum Handeln.

Das Skript loggt für jede Antwort eines DNS-Servers die aufgelöste URL und dazugehörige IP-Adresse (Z. 50).

Parallel dazu wird periodisch die Hosts-Datei eines Unix-Systems abgefragt (Z. 63).

Im Log vermerkt sind ausgehende Verbindungen, deren Zieladressen vorher nicht durch einen DNS-Server oder die Hosts-Datei aufgelöst wurden (Z. 68).

Quelltext 5.5: Überprüfung der Verbindungsziele eines Endgerätes

```

45 event DNS::log_dns(rec:DNS::Info){
46     local query : string = rec$query;
47     local answer : vector of string = rec$answers;
48     local answer_address = to_addr(answer[1]);
49     resolved_addresses [answer_address] = query;
50 }
51
52
53
54
55
56
57
58
59 function query_hosts_file(){
60     local query_event = query_result;
61     local _schedule = 30 secs;
62     local hosts_file_query = ZeekAgent::query([$sql_stmt=str_stmt_hosts,
63         $event=query_event, $schedule=_schedule]);
64 }
65
66 event check_resolve_table(c : connection){
67     local destination_ip = c$resp_h;
68     if(destination_ip !in resolved_addresses && destination_ip !in dns_server){
69         local rec: DNStest::Info = [$ts=current_time(), $id=c$id,
70             $notice="Connection without Resolve!"];
71         c$dnstest = rec;
72         Log::write(DNStest::LOG, rec);
73     }
74 }

```

6 Evaluation

In diesem Absatz soll begutachtet werden, inwiefern sich die im Kapitel 5 vorgestellten kontextsensitiven IDS-Skripte auf die im Kapitel 4 priorisierten Leistungsmetriken eines IDS auswirken und inwiefern eine Verbesserung dieser Metriken möglich ist. Ziel ist also zu evaluieren, ob sich die Leistung eines IDS verbessern lässt, wenn in Zugriffsentscheidungen Kontextinformationen miteinbezogen werden und welche Informationen sich besonders gut dazu eignen. Dazu wurde zuerst analysiert, ob und wie vorhandener Kontext im Bezug auf NSM kategorisiert werden kann und welche Kontextinformationen sich zur Anpassung der Arbeitsabläufe eines IDS eignen. Diese sollten zur besseren Übersicht in einer Taxonomie in Beziehung zueinander gesetzt werden. Zusätzlich wurde erörtert wie und wo diese Informationen zu beschaffen sind. Bewertet werden sollte dabei, inwiefern es Möglichkeiten der Anpassung an sich verändernde Bedingungen gibt und Kontextinformationen zur Verbesserung der Leistung eines IDS beitragen. Das Hauptaugenmerk liegt dabei darauf, falsch-positive Meldungen zu verringern und die Interpretierbarkeit der Logs zu erhöhen, indem zusätzliche Kontextinformationen in die Entscheidungsfindung des IDS miteinbezogen werden. Dazu wurde untersucht, wie gut sich die für den jeweiligen Anwendungsfall benötigten Informationen sammeln und verarbeiten lassen. Anhand dessen soll eine Einschätzung darüber gegeben werden, welche Kontextkategorien sich zur Verbesserung der IDS-Metriken eignen. Dabei wurden Kontextinformationen aus verschiedenen Einzelkategorien der im Kapitel 4 vorgestellten Taxonomie kombiniert, um im jeweiligen Anwendungsfall eine Verbesserung der IDS-Eigenschaften zu erreichen. Die Nützlichkeit der Kontextinformationen hängt dabei von der Verfügbarkeit und Qualität der Informationen sowie davon wie komplex das dazugehörige Skript ist ab. Die Verfügbarkeit wird daran bemessen wie anspruchsvoll die Informationen zu beschaffen waren. Die Qualität wird von der Höhe des Aufwandes die Informationen in eine Form zu bringen, die eine Verarbeitung zulässt.

6.1 Bedeutung von Kontextsensitivität

Um die im Kapitel 4 besprochenen Nachteile auszugleichen, bietet es sich an, die bereits verfügbaren, gesammelten und aufbereiteten Kontextinformationen zu verwenden. Dies ermöglicht, die schon existenten Erkennungsmechaniken so anzupassen, dass sie eine größere Menge von Ereignissen abdecken oder schon definierte Ereignisse genauer zu spezifizieren und so weniger Meldungen zu generieren, ohne dabei das Verhalten des Netzwerkverkehrs dauerhaft neu bewerten zu müssen.

6.2 IDS-Implementierung

Zeek bot im Vergleich zu anderen IDS die beste Software bzw. Umgebung für die Entwicklung neuer kontextsensitiver Erkennungs- oder Verarbeitungstechniken. Es kann für die kontinuierliche Überwachung von Netzwerken mit hohem Durchsatz verwendet werden. Die Skripting-Umgebung ist in einer speichersicheren Sprache erweiterbar, die auf die Verarbeitung von Netzwerkdaten spezialisiert ist. Im Gegensatz zu anderen Tools ist es nicht auf ein einziges Paradigma für die Netzüberwachung limitiert. Aber auch Zeek in Kombination mit Zeek-Agent ist nicht perfekt. Die Vor- und Nachteile der in dieser Arbeit verwendeten Toolchain werden in den nachfolgenden Abschnitten an gegebener Stelle diskutiert.

6.3 Vergleich der Kontextkategorien

Ähnlich wie in den im Bereich 4 verwendeten Grafiken erfolgt auch die Bewertung der Unterteilung im Hinblick auf Verfügbarkeit und Qualität. Auf die im Teilbereich 5 implementierten Rubriken und dabei aufgetretene Probleme wird jeweils noch einmal separat eingegangen. Alle weiteren im Kapitel 4 aufgestellten Kategorien werden anhand der Erkenntnisse, die während des Implementationsprozesses entstanden sind, ausschließlich in der Grafik 6.1 bewertet. Dabei erfolgt auch eine Veranschaulichung davon welche Kategorie zur Erfüllung welches Ziels geeignet ist. Dargestellt wird weiterhin, ob eine Kategorie zur Verbesserung der Metriken für sich allein, oder nur in Kombination mit anderen Kategorien beitragen kann.

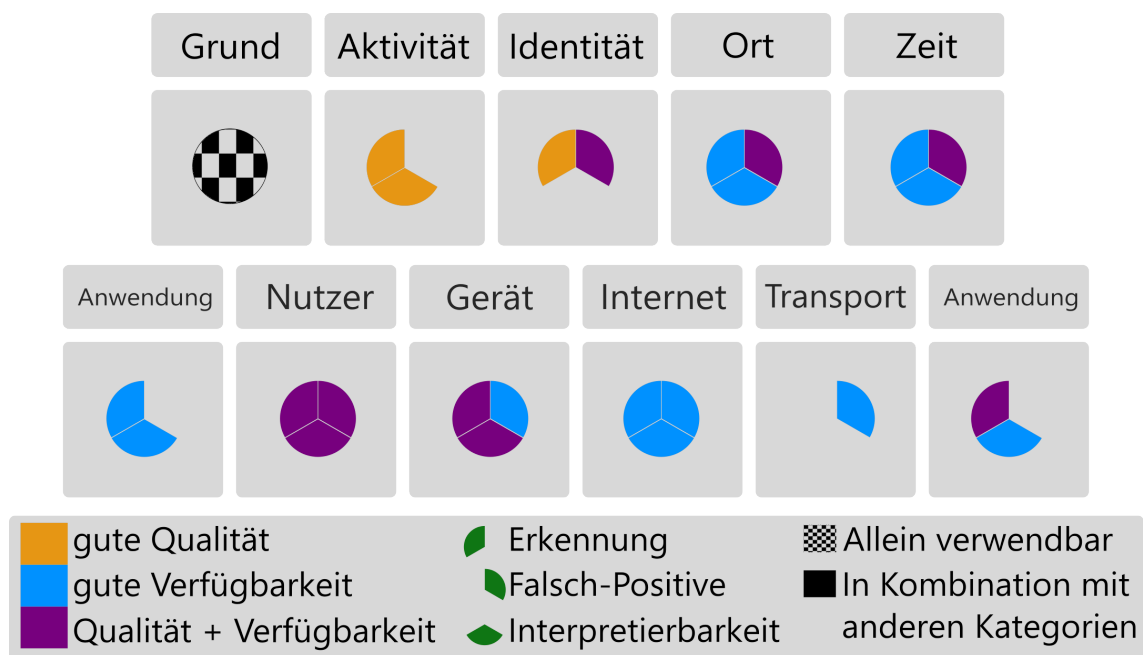


Abbildung 6.1: Evaluation der Taxonomie

6.3.1 Verfügbarkeit der Informationen

Die Informationen, die im Kapitel 4 beispielhaft als gegeben vorausgesetzt werden, sind in der Praxis unterschiedlich schwer zu akquirieren. Es ist beispielsweise um ein Vielfaches einfacher, korrekt die aktuelle Uhrzeit zu ermitteln als den Grund dafür, dass ein Nutzer eine Aktion ausführt, zu bestimmen. Auch sind einige Informationen für die Funktionalität des

Netzwerkes und der Kommunikation zwingend notwendig, andere hingegen sind optional oder auf bestimmten Systemen nicht vorhanden.

Anwendungsfälle

Die Informationen, die mit Hilfe des Zeek-Agent direkt vom Host abgefragt werden, sind, da sie ohnehin zum korrekten Betrieb des Systems in der Praxis gebraucht werden, stets verfügbar. Einzig die Zeit, die eine Zeek-Agent-Instanz für eine Antwort benötigt, hat sich bei komplexeren Anfragen oder zu langsamer Hardware als problematisch erwiesen. Es ist also weniger ein Problem, dass Informationen generell nicht vorhanden sind, sondern lediglich nicht ausreichend schnell zur Verfügung gestellt werden können. Im Fall der verwendeten Ports (5.4) führte die Komplexität der an den Zeek-Agent gesendeten Anfrage dazu, dass eine Antwort mitunter erst eintraf, nachdem schon zum Hostsystem zugehöriger Netzwerkverkehr bearbeitet wurde. Dies kann vor allem in Netzwerken mit einem hohen Durchsatz und langsam antwortenden Hostsystemen zu fälschlicherweise geblocktem Verkehr führen. Zu einem gewissen Grad lässt sich diese Problematik mit festem Scheduling einzelner Events umgehen. Das Scheduling verschiedener Ereignisse mit teilweise asynchronen Antworten wirft allerdings neue Probleme auf, welche an dieser Stelle nicht näher erläutert werden. Es empfiehlt sich deshalb, statt eines großen komplexen Skripts mehrere kleinere Skripte zu verwenden, um Leistungseinbrüche zu vermeiden. Dieser modulare Ansatz hat außerdem den Vorteil, dass sich die Menge an Ergebnissen, die miteinander in Bezug gesetzt werden können, vergrößert. So können über einen längeren Zeitraum oder mehrere Geräte verteilte Angriffe noch besser erkannt werden, als wenn lediglich Kontextsensitivität allein verwendet wird.

6.3.2 Qualität der Informationen

Auch die Qualität der Informationen ist in der Praxis limitiert. Genauigkeit und Korrektheit der Angaben sind teilweise stark unterschiedlich. Auch die Vertrauenswürdigkeit und Aktualität der Daten variiert. Es kann im Allgemeinen davon ausgegangen werden, dass Informationen, die von einem Host, der sich im eigenen Netz befindet, stammen, nicht auf dem Weg zum IDS manipuliert werden. Ein Problem stellt dabei allerdings die Tatsache dar, dass ein Hostsystem und die Daten, die es an das IDS sendet, kompromittiert sein können. Mit einer zunehmenden Menge an Kriterien, die in Betracht gezogen werden, um die Unauffälligkeit von Kommunikation, Verhalten und Zustand zu bemessen, steigt auch die Anzahl von Datenpunkten, die gleichzeitig von einem Angreifer verfälscht werden müssen. Durch einen stetigen Anstieg an gesammelten Daten wird die Dauer, für die ein Angreifer unbemerkt ein System kontrollieren kann und damit das Schadenpotenzial verringert, zumindest wenn angenommen wird, dass alle Metriken gleich schwer vorzutäuschen sind. Selbst wenn das Hostsystem also keine korrekten Daten mehr zur Verfügung stellt, kann dies angesichts ausreichend vieler anderer Daten ausgeglichen werden und sollte dies nicht möglich sein, so wird der Angreifer potenziell zumindest schneller als ohne die Verwendung dieser Zusatzinformationen erkannt. Netzwerkverkehr, der von außerhalb des Netzwerkes stammt, ist mit einer höheren Wahrscheinlichkeit verfälscht, gerade wenn die verwendeten Kommunikationsprotokolle und deren Struktur keine inhärenten Möglichkeiten bieten, Informationen oder die Seriosität des Senders zu prüfen. Auch gibt es hier weniger Möglichkeiten, zusätzliche Informationen zu gewinnen.

6.3.3 Anwendungsfälle

Der momentan limitierende Faktor im Fall der Verwendung von Hostinformationen sind die mit Zeek-Agent Kategorien von abfragbaren Daten. In einer früheren Version des Zeek-Agent bestand die Option, externe Tools wie `osquery`¹ welches zum gegenwärtigen Zeitpunkt für Linux, statt der 7 SQL-Tabellen des Zeek-Agents, 154 Tabellen zur Verfügung stellt², anzufragen und damit eine wesentlich größere Menge an Informationen als im momentanen Entwicklungsstand verwenden zu können. Dieser Ansatz wurde allerdings von den Entwicklern aufgrund der zu hohen Komplexität verworfen. Das limitiert in vielen Fällen die abdeckbaren Anwendungsfälle. Nichtsdestotrotz deckt Zeek-Agent in Kombination mit Zeek die im Kapitel 4 vorgeschlagenen Kategorien in ihren Grundzügen sofern möglich, weitestgehend ab.

6.3.4 Vertrauenswürdigkeit der Informationsquellen

Bestenfalls stammen Informationen von einer allgemein vertrauenswürdigen Quelle. Außerhalb des Netzwerkes kann das, wie etwa im Skript 5.5, ein autoritativen DNS-Nameserver sein. Innerhalb des Systems sind das zum Beispiel Informationen, auf die ein Endnutzer ohne erweiterte Rechte nur begrenzten oder keinen Einfluss ausüben kann oder das meist ohne bewusstes Zutun, wie beispielsweise im Skript 5.4 die Wahl des Ports geschieht. Denkbar wären auch Geräte, die begrenzt beeinflussbar sind oder auf die ausschließlich bestimmte Personen physischen Zugriff haben. Ein so ermittelter Datenpunkt ist ein guter Indikator, um potenziell auffälligen Netzwerkverkehr zu identifizieren.

6.4 Leistungsverbesserung

Die drei Anwendungsbeispiele haben demonstriert, dass Kontext sowohl die Erkennung von Angriffen und die Rate an Falsch-positiven als auch die Interpretierbarkeit des Geschehens im Netzwerk verbessert. Dabei hat sich gezeigt, dass keine Kategorie allein alle Leistungsmetriken simultan zufriedenstellend abdecken und gleichzeitig in ausreichendem Maß zur Verfügung steht. Vielmehr ist es mitunter so das je besser eine Kategorie die Ziele erfüllen, umso schwerer fällt es, sie zu akquirieren. Die Kategorie Grund deckt beispielsweise alle Metriken ab, wenn man die dazu benötigten Informationen zur Verfügung stehen, allerdings ist in der Praxis quasi unmöglich, diese Informationen zu erhalten. Zusätzlich ergibt sich noch die Verbesserung der Vertrauenswürdigkeit der Kontextinformationen, auch wenn diese beim Design der Taxonomie und dem Entwurf der Anwendungsfälle nicht beachtet wurde, da eine Analyse und Bewertung der Vertrauenswürdigkeit der Kontextinformationsquellen nicht Fokus dieser Arbeit war. Wie schon bei der Qualität der Informationen diskutiert, kann festgehalten werden, das sich der Aufwand, ein gewisses Maß an Vertrauen bei beispielsweise einer Kontrollinstanz wie einem IDS in eine bestimmte Entität zu erreichen für diese Entität mit der Menge und der Qualität der dafür benötigten Informationen erhöht. Durch die mit den neuen Informationen für das IDS eröffneten Möglichkeiten wird das Vortäuschen falscher Tatsachen und der damit verbundene Vertrauensgewinn in Entitäten für Angreifer zusätzlich erschwert.

¹<https://osquery.io/>

²<https://osquery.io/schema/5.5.1/>

7 Zusammenfassung und Ausblick

Es gibt eine große Menge an möglichen Ansätzen Kontext einzuteilen. Einige davon wurden in dieser Arbeit vorgestellt, zwei selbst in Kapitel 4 konstruiert, in Kapitel 5 exemplarisch getestet und anschließend in Kapitel 6 ausgewertet. Es hat sich dabei gezeigt, dass Kategorien bis auf wenige Ausnahmen erst in Kombination mit anderen Kategorien ihr volles Potenzial entfalten können und selten allein funktionieren. Zeek hat sich dabei aufgrund seiner vielfältigen Optionen als ein sehr gutes Werkzeug zum Testen des vorgeschlagenen Designs herausgestellt. Die Möglichkeiten, welche die gewählte Implementation bietet, wenn erst einmal verstanden wurde, wie sie zu nutzen ist, sind fast unbegrenzt. Zeek-Agent hingegen hat sich als weniger mächtig als zunächst vermutet herausgestellt. Positiv zu erwähnen ist die hervorragende Einbindung in Zeek. Negativ ist allerdings die im Vergleich zu anderen solchen Werkzeugen geringe Anzahl an abfragbaren Informationen. Hier bietet es sich also an, eventuell zusätzliche, in Kapitel 4 vorgestellte Mittel zu verwenden. Zusammengefasst lässt sich aber sagen, dass mehr Kontextinformationen in den allermeisten Fällen die Leistung verbessern.

An anderer Stelle geklärt werden muss, wie damit umgegangen werden soll, wenn keine einzige der zusätzlichen Datenquellen ausreichend vertrauenswürdig ist. Dazu sollte getestet werden, welche Grundvoraussetzungen eine glaubwürdige Informationsquelle erfüllen sollte.

Zukünftig zu untersucht werden sollte auch, welche Probleme durch die im Prozess dieser Arbeit erzeugte und zusammengefügte Menge an Daten entstehen und wie diese zu lösen sind. Voraussichtlich ist dabei vordergründig zu beantworten, wie die Integrität der Informationen durch mutwillige Angriffe und fehlerhaft konfigurierte Geräte beeinträchtigt wird.

Ein weiterer in dieser Arbeit nicht betrachteter Aspekt sind Prozesse, die in ihrem Lebenszyklus keine Daten auf die Festplatte schreiben oder anderweitig nachweisbar Spuren auf den Geräten, auf denen sie ausgeführt werden, hinterlassen. Es wurde sich nur mit Vorgängen beschäftigt, die nachweislich einlesbare Daten erzeugen oder für die Gewährleistung ihrer Funktionalität zumindest benötigen. Netzwerkverkehr von Schadsoftware, die beispielsweise ausschließlich im Arbeitsspeicher eines Systems arbeitet, kann deshalb nicht mit zusätzlichen Kontextinformationen kombiniert werden. Die Erkennung solcher Angriffe kann nur sehr begrenzt von den Erkenntnissen dieser Arbeit profitieren. Hier wäre eine zukünftige Betrachtung der Erkennungsmöglichkeiten also dringend notwendig, da gerade diese Art von Software ein zunehmendes Problem im Bereich der IT-Sicherheit darstellt und es nach Wissensstand des Autors bisher noch wenige ausgereifte Ansätze zur Bekämpfung gibt.

Literatur

- [1] *About Zeek — Book of Zeek (git/master)*. URL: <https://docs.zeek.org/en/master/about.html> (besucht am 21.09.2022).
- [2] Gregory D. Abowd u. a. "Towards a Better Understanding of Context and Context-Awareness". In: *Handheld and Ubiquitous Computing*. Hrsg. von Hans-W. Gellersen. Bearb. von Gerhard Goos, Juris Hartmanis und Jan van Leeuwen. Bd. 1707. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, S. 304–307. ISBN: 978-3-540-66550-2 978-3-540-48157-7. DOI: 10.1007/3-540-48157-5_29.
- [3] Jose Aguilar, Marxjhony Jerez und Tania Rodríguez. "CAMEnto: Context awareness meta ontology modeling". en. In: *Applied Computing and Informatics* 14.2 (Juli 2018), S. 202–213. ISSN: 22108327. DOI: 10.1016/j.aci.2017.08.001.
- [4] Unai Alegre, Juan Carlos Augusto und Tony Clark. "Engineering context-aware systems and applications: A survey". en. In: *Journal of Systems and Software* 117 (Juli 2016), S. 55–83. ISSN: 01641212. DOI: 10.1016/j.jss.2016.02.010.
- [5] Mary Bazire und Patrick Brézillon. *Understanding Context Before Using It*. en. Hrsg. von David Hutchison u. a. Bd. 3554. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, S. 29–40. ISBN: 978-3-540-26924-3 978-3-540-31890-3. DOI: 10.1007/11508373_3.
- [6] Suan Khai Chong u. a. "Context-aware sensors and data muffling". In: *Context awareness for self-managing systems (devices, applications and networks) proceeding* (2007), S. 103–117.
- [7] John D Day und Hubert Zimmermann. "The OSI reference model". In: *Proceedings of the IEEE* 71.12 (1983), S. 1334–1340.
- [8] Anind K Dey. "Understanding and using context". In: *Personal and ubiquitous computing* 5.1 (2001), S. 4–7.
- [9] Ibrahim Ghafir, Jakub Svoboda und Vaclav Prenosil. "Network Monitoring Approaches An Overview". en. In: *Third International Conference on Advances in Computing, Communication and Information Technology- CCIT 2015*. Institute of Research Engineers und Doctors, Mai 2015, S. 118–123. ISBN: 978-1-63248-061-3. DOI: 10.15224/978-1-63248-061-3-72.
- [10] Saul Greenberg. "Context as a dynamic construct". In: *Human-Computer Interaction* 16.2-4 (2001), S. 257–268.

- [11] Karen Henricksen. "A framework for context-aware pervasive computing applications". School of Information Technology und Electrical Engineering, The University of Queensland, 2003. URL: <http://henricksen.id.au/publications/phd-thesis.pdf>.
- [12] Anita K Jones und Robert S Sielken. "Computer system intrusion detection: A survey". In: *Computer Science Technical Report* (2000), S. 1–25.
- [13] A. S. M. Kayes, Jun Han und Alan Colman. "ICAF: A Context-Aware Framework for Access Control". In: ACISP'12. Wollongong, NSW, Australia: Springer-Verlag, 2012. ISBN: 9783642314476. DOI: 10.1007/978-3-642-31448-3_34.
- [14] A. S. M. Kayes u. a. "A Survey of Context-Aware Access Control Mechanisms for Cloud and Fog Networks: Taxonomy and Open Research Issues". en. In: *Sensors* 20.9 (2020), S. 2464. ISSN: 1424-8220. DOI: 10.3390/s20092464.
- [15] A.S.M. Kayes, Jun Han und Alan Colman. "An ontological framework for situation-aware access control of software services". en. In: *Information Systems* 53 (2015), S. 253–277. ISSN: 03064379. DOI: 10.1016/j.is.2015.03.011.
- [16] Peter Mell u. a. *An Overview of Issues in Testing Intrusion Detection Systems*. en. 2003. DOI: <https://doi.org/10.6028/NIST.IR.7007>.
- [17] Aleksandar Milenkoski u. a. "Evaluating Computer Intrusion Detection Systems: A Survey of Common Practices". en. In: *ACM Computing Surveys* 48.1 (2015), S. 1–41. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/2808691.
- [18] Charith Perera u. a. "Context Aware Computing for The Internet of Things: A Survey". In: *IEEE Communications Surveys & Tutorials* 16.1 (2014), S. 414–454. ISSN: 1553-877X. DOI: 10.1109/SURV.2013.042313.00197.
- [19] Lexi Pimenidis, Benedikt Westermann und Volker Wetzelaer. "A context aware network-IDS". In: *Proceedings of The 13th Nordic Workshop on Secure IT Systems NordSec 2008*. 2008, S. 3.
- [20] Karen Scarfone, Peter Mell u. a. "Guide to intrusion detection and prevention systems (idps)". In: *NIST special publication* 800.2007 (2007), S. 94.
- [21] Bill Schilit, Norman Adams und Roy Want. "Context-aware computing applications". In: *1994 first workshop on mobile computing systems and applications*. IEEE. 1994, S. 85–90.
- [22] Bruce Schneier. "Managed Security Monitoring: Network Security for the 21st Century". In: *Computers & Security* 20.6 (Sep. 2001), S. 491–503. ISSN: 01674048. DOI: 10.1016/S0167-4048(01)00607-1.
- [23] Arthur H Van Bunningen, Ling Feng und Peter MG Apers. "Context for ubiquitous data management". In: *International Workshop on Ubiquitous Data Management*. IEEE. 2005, S. 17–24.
- [24] Wei Liu, Xue Li und Daoli Huang. "A survey on context awareness". In: *2011 International Conference on Computer Science and Service System (CSSS)*. 2011 International Conference on Computer Science and Service System (CSSS). Nanjing, China: IEEE, Juni 2011, S. 144–147. ISBN: 978-1-4244-9762-1. DOI: 10.1109/CSSS.2011.5972040.
- [25] J. Wolfgang Kaltz, Jürgen Ziegler und Steffen Lohmann. "Context-aware Web Engineering: Modeling and Applications". In: *Revue d'intelligence artificielle* 19.3 (1. Juni 2005), S. 439–458. ISSN: 0992499X. DOI: 10.3166/ria.19.439-458.
- [26] *zeek-agent-v2: Open source endpoint agent providing host information to Zeek. [v2]*. URL: <https://github.com/zeek/zeek-agent-v2> (besucht am 21. 09. 2022).

- [27] Andreas Zimmermann, Andreas Lorenz und Reinhard Oppermann. "An Operational Definition of Context". In: *Modeling and Using Context*. Hrsg. von Boicho Kokinov u. a. Bd. 4635. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, S. 558–571. ISBN: 978-3-540-74254-8. DOI: 10.1007/978-3-540-74255-5_42.