PJ Intelligent Cybersecurity Applications

Implementation of an IT-Environment Simulator for a Reinforcement Learning Agent

Jonathan Ackerschewski, Markus Bartels, Linus Schacht

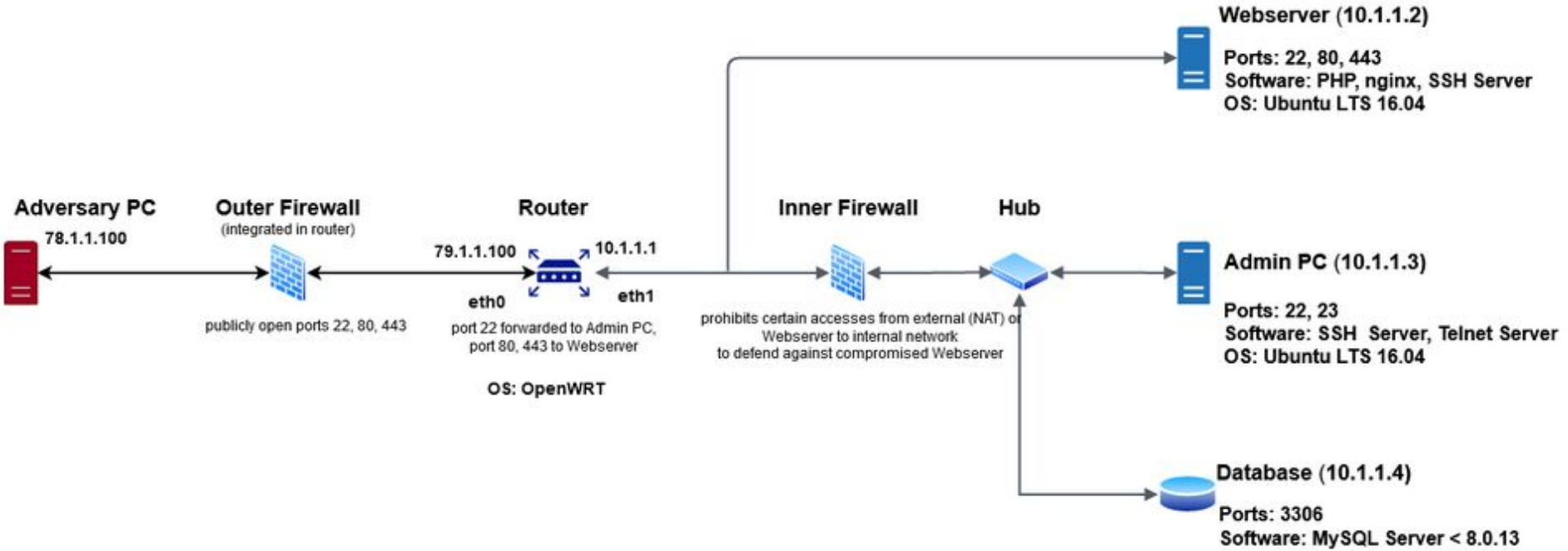
Aufgabenstellung

- Entwurf eines Modells und Simulation eines realistischen Angriffs auf ein kleines Netzwerk
- Nützlich für: Threat Assessment, Training, Forecasting
- Unser Ansatz: Reinforcement Learning mit Q-Learning Algorithmus (Lernen der optimalen Policy für einen MDP)

Inhaltsverzeichnis

1. Szenario
2. Policy Evaluation
3. Learning Evaluation
4. Fazit
5. Stand der Technik

Netzwerk, Hosts und Services



https://gitlab.tu-berlin.de/j.ackerschewski/ICA-2/wikis/Network-Topology

Angreifer Aktionen (nach Mitre ATT&CK [1])

Aktionen:

- Active Scanning (IP & Schwachstellen)
- Exploit Public Facing Application (z.B. RCE)
- Exploit for Client Execution (z.B. Code Injection)
- Exploit for Privilege Escalation
- Valid Account
- Data from Local System
- Software Discovery
- Man-in-the-Middle
- ~~Create Account~~

Motivation:

- Abbilden des Angriffs-Lebenzyklus (Reconaissance, Initial Access, Privilege Escalation, ...)
- Darunter auch Lateral Movement
- Angreifer hat die Wahl: Schwachstelle suchen und ausnutzen, Authentifizierung umgehen durch Finden eines geheimen Schlüssels, Netzwerktraffic belauschen...
- Aktionen haben verschiedenen **Nutzen**, aber auch verschiedene **Voraussetzungen**

[1] https://attack.mitre.org/tactics/enterprise/

Das Ziel des Angreifers vs. Gegenmaßnahmen

- Root Rechte auf allen Hosts
- ~~Persistenz erlangt auf den Hosts Admin PC sowie Datenbank~~
- Auslesen der geheimen Daten in der Datenbank
- Gesamtes Netzwerk gescannt
- Zeroday-Exploit nur wenn nötig benutzen

- Firewalls
- Honeypots

Policy Evaluation - Vorgehen

- Vorüberlegungen
- Performance zufälliger Aktionen (mit und ohne Eigentransitionen)
  - Vermutung: ohne Eigentransitionen besser, da durch mehrfache Aktionen in unserem Modell kein Informationsgewinn
- Auswertungen zu Utility Funktion
  - Inkl. Prüfen, ob Pfad der höchsten Utility = beste Policy
- Auswertungen zu verschiedenen Honeypot-Szenarien
  - Erwartung: Alternativer Weg wird gefunden
- Vorstellen beste gelernte Policy

Policy Evaluation - Was ist objektiv gut...?

... für ein schnelles Erreichen des Ziels (**Aktionsanzahl**!)

- Keine Software Discovery
- Falls möglich, Valid Account > Exploit (Annahme: weniger auffällig)
- Data From Local System üblicherweise besser wenn bereits Root Rechte erlangt
- Innere Firewall blockiert Vuln-Scanning & Exploit von WS aus auf DB: besser MitM oder über AdminPC
- Keine Aktionen vom Angreifer Host aus, sobald bereits ein interner Host übernommen wurde
- Umgehen von **Honeypot & Zeroday**, falls aktiv

Policy Evaluation - Optimale Parameter

Learning Parameter:

Learning Rate (Start|End|Max|Slope): 0.4 | 0.05 | 10 | 1.0
Epsilon (Start|End|Slope): 0.25 | 0.0 | 4.0
Discount Factor: 1.0
Error: 1e-09
Ne: 5
R-Plus: 20.0
Iterations: 0
Initial State Iterations: 500000 (500k)

“Normale” Utility Funktion:

- **Kosten je Aktion**: Scans & Valid Account niedrig (0.1), Exploit hoch (0.3)
- Belohnung für Erreichen **Root Zugriff**: 0.5
- Belohnung für internes Scannen: 0.25 | 0.15
- Kosten für Nutzen des Adversary Hosts nach initialem internen Netzzugriff: 0.5
- **Zeroday Bestrafung**: 3.0
- **Honeypot Bestrafung / Zielzustand Belohnung**: 5.0

Alle jeweils nicht spezifizierten Parameter gelten für die folgenden Folien

Policy Evaluation - Zufällige Aktion mit Eigentransitionen

ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: VALID_ACCOUNTS_VULN
	Target: WEBSERVER	Action: EXPLOIT_FOR_CLIENT_EXECUTION
WEBSERVER	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: MAN_IN_THE_MIDDLE
	Target: ADMINPC	Action: ACTIVE_SCAN_IP_PORT
ADMINPC	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
WEBSERVER	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
ADMINPC	Target: WEBSERVER	Action: EXPLOIT_FOR_PRIVILEGE_ESCALATION
	Target: DATABASE	Action: VALID_ACCOUNTS_CRED
	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM

Iterationen: 500k
Minimum transitions: 14
Maximum transitions: 666
Mean transitions: **98.28**
Median transitions: 85.0
Mode of transitions: 62
Standard deviation transitions: **52.33**
Shortest Policy length was found 13 times

Policy Evaluation - Zufällige Aktion ohne Eigentransitionen

ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: VALID_ACCOUNTS_VULN
ADMINPC	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
	Target: DATABASE	Action: VALID_ACCOUNTS_VULN
DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM
ADMINPC	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: EXPLOIT_FOR_CLIENT_EXECUTION
DATABASE	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: EXPLOIT_FOR_PRIVILEGE_ESCALATION

Iterations: 500k (MDP)
Shortest Policy Size **12**: Reward -2.65
Most Reward Policy Size 16: Reward **5.70**
Maximum transitions: 23
Mean transitions: 19.22
Median transitions: **19**
Mode transitions: 19
SD transitions: 1.37
Maximum reward: 5.70
Mean reward: **1.66**
SD reward: 0.99

Policy Evaluation - Greedy (max. Utility)

ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: EXPLOIT_PUBLIC_FACING_APPLICATION
	Target: WEBSERVER	Action: EXPLOIT_FOR_PRIVILEGE_ESCALATION
WEBSERVER	Target: ADMINPC	Action: ACTIVE_SCAN_IP_PORT
	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: SOFTWARE_DISCOVERY
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: DATA_FROM_LOCAL_SYSTEM
	Target: ADMINPC	Action: VALID_ACCOUNTS_CRED
ADMINPC	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
	Target: DATABASE	Action: VALID_ACCOUNTS_VULN
DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM

Aktionen: **17**

Reward: **5.30**

Policy Evaluation - Utility Funktion ohne Scan Vorteil

Active Host: ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: EXPLOIT_FOR_CLIENT_EXECUTION
Active Host: WEBSERVER	Target: WEBSERVER	Action: DATA_FROM_LOCAL_SYSTEM
	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: VALID_ACCOUNTS_CRED
	Target: WEBSERVER	Action: DATA_FROM_LOCAL_SYSTEM
	Target: ADMINPC	Action: VALID_ACCOUNTS_CRED
Active Host: ADMINPC	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: MAN_IN_THE_MIDDLE
Active Host: WEBSERVER	Target: DATABASE	Action: VALID_ACCOUNTS_CRED
Active Host: DATABASE	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: ACTIVE_SCAN_IP_PORT
Active Host: WEBSERVER	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
Active Host: DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM

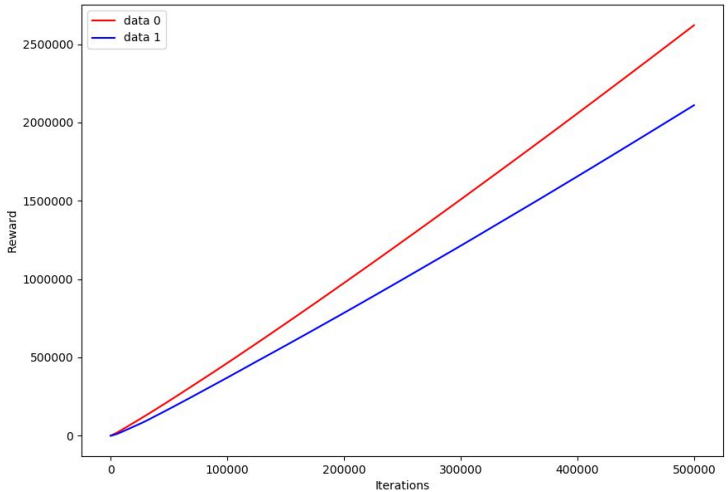
Aktionen: 16
Reward: 4.6

Beobachtung: Keine unnötigen/schlechten Aktionen & Priorität auf Root Access
Aber: nachgezogenes Scannen, oft Wechseln des Hosts

Policy Evaluation - Utility Funktion mit Scan Vorteil

ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: EXPLOIT_FOR_CLIENT_EXECUTION
WEBSERVER	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: DATA_FROM_LOCAL_SYSTEM
	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: VALID_ACCOUNTS_CRED
	Target: WEBSERVER	Action: DATA_FROM_LOCAL_SYSTEM
	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: VALID_ACCOUNTS_CRED
ADMINPC	Target: ADMIN	Action: MAN_IN_THE_MIDDLE
	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
WEBSERVER	Target: DATABASE	Action: VALID_ACCOUNTS_CRED
DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM

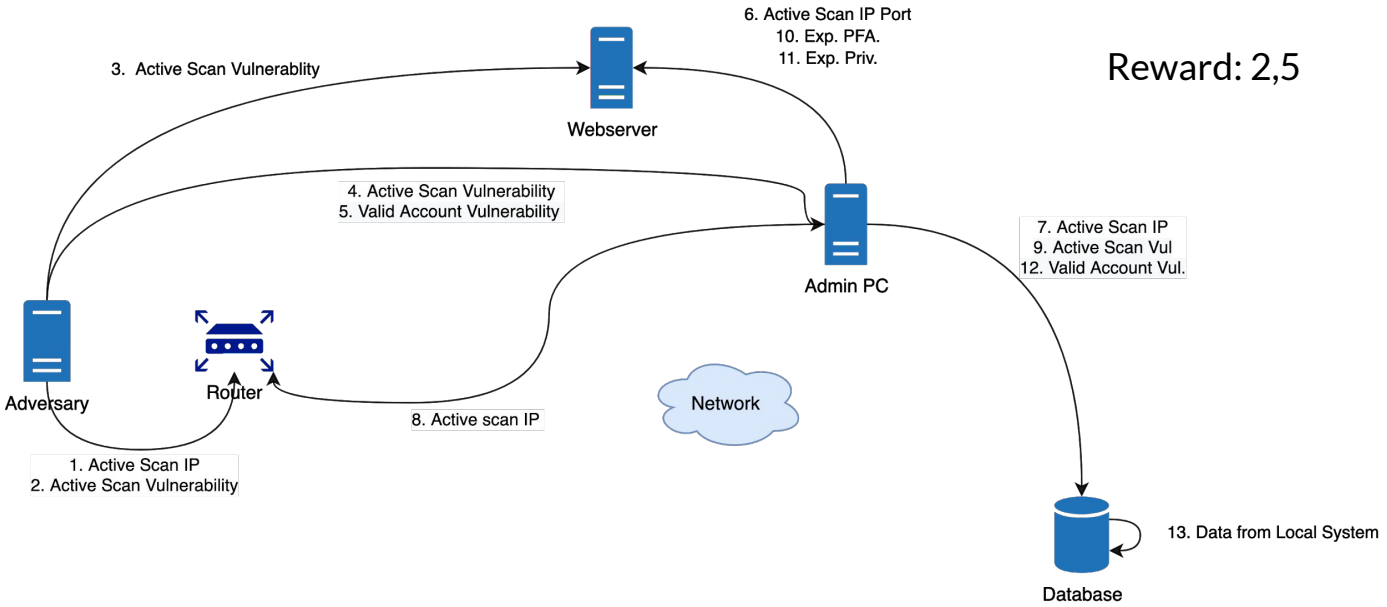
Vergleich Reward



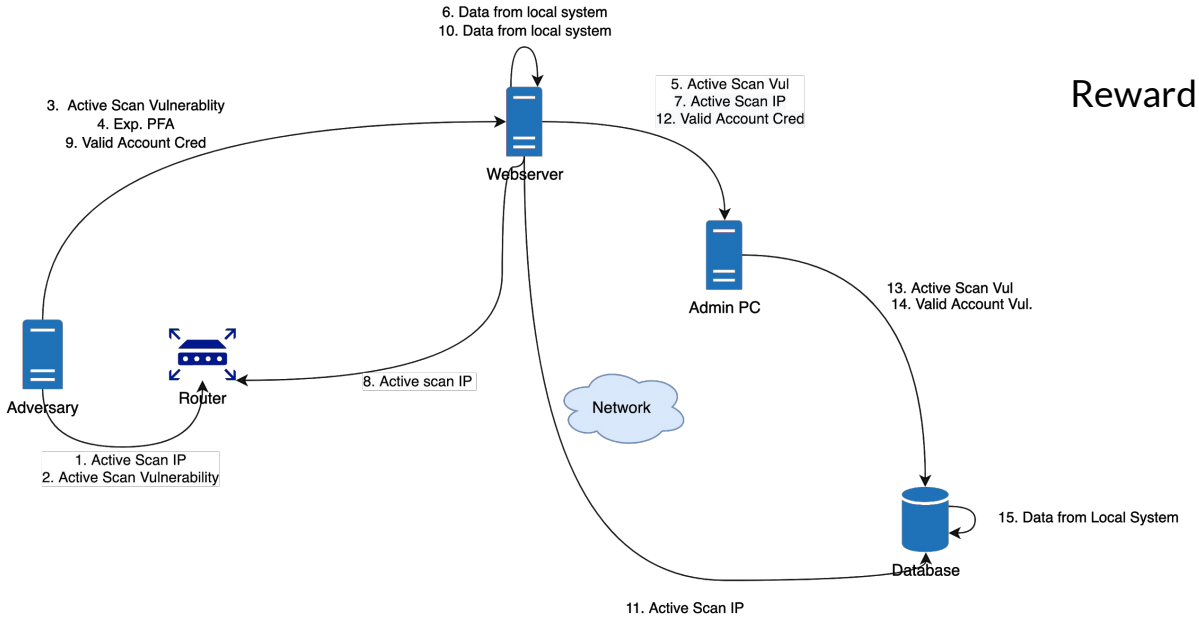
Keine unnötigen Aktionen, Balance zw. Root Zugriff & Scanning
Gleiche Anzahl Aktionen (16), Reward 5.7

Rot: Scan Vorteil Blau: ohne

Policy Evaluation Honeypot: Admin PC Cred. vom Webserver

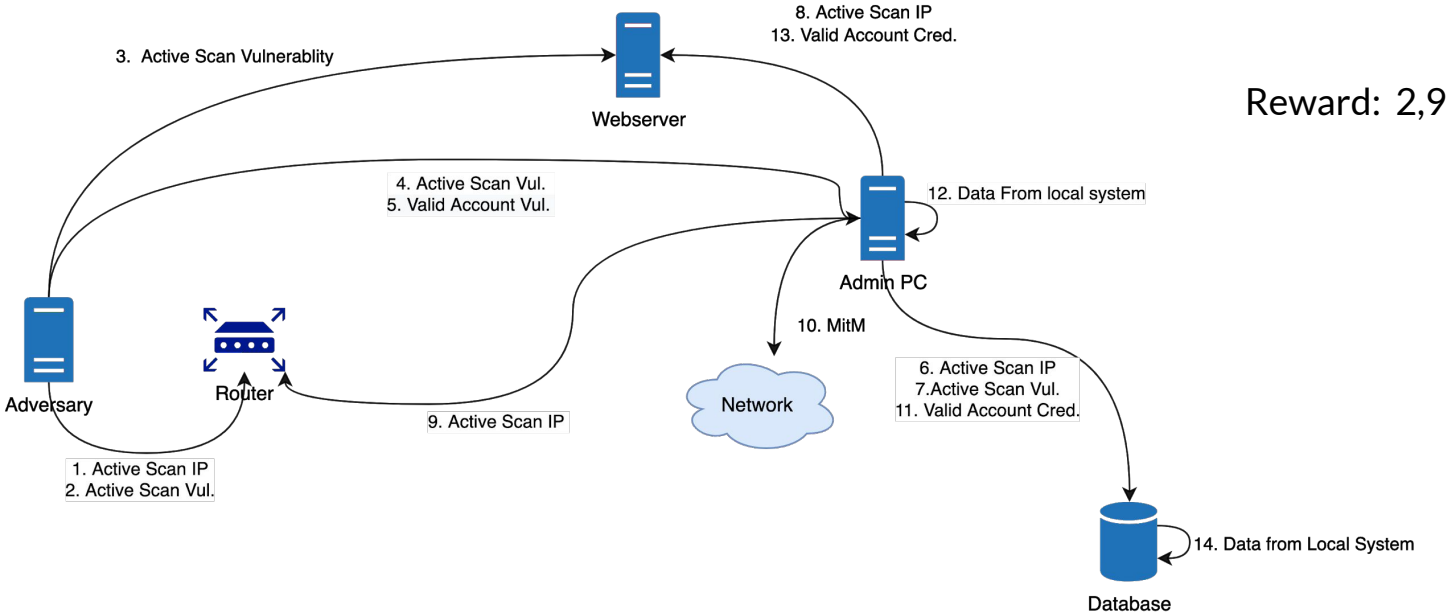


Policy Evaluation Honeypot: Credentials from MitM



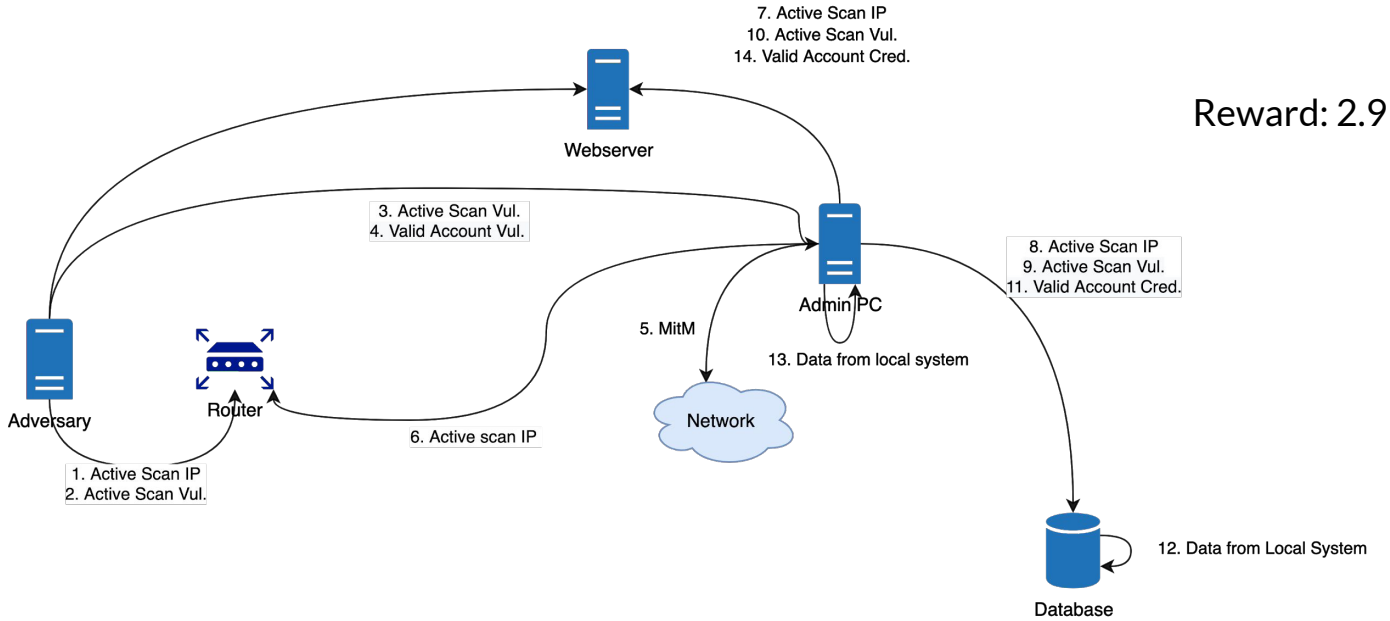
Reward: 5.7

Policy Evaluation Honeypot: Exploits Webserver



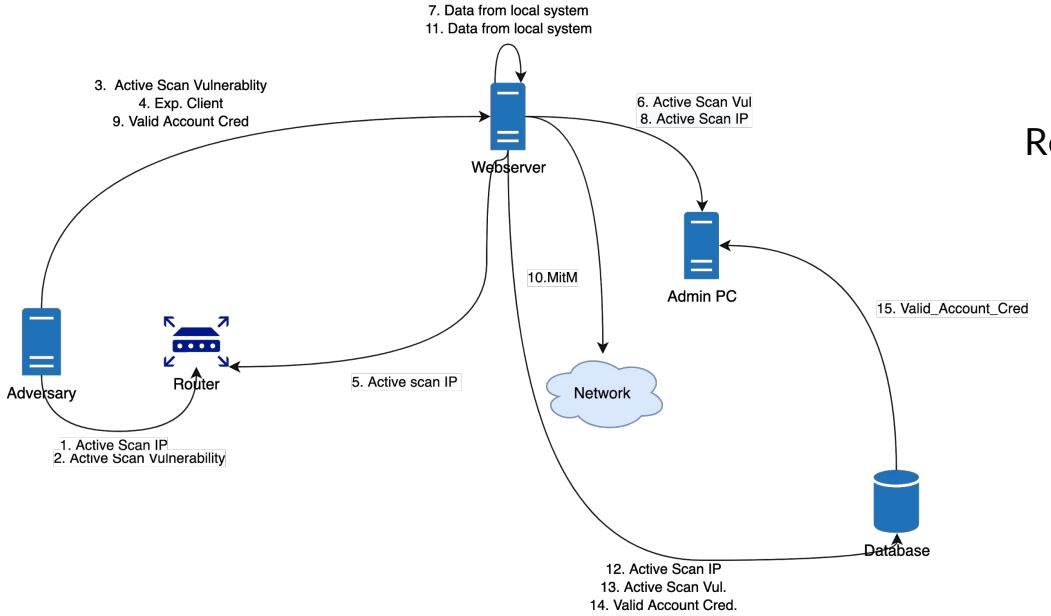
Reward: 2,9

Policy Evaluation Honeypot: Privilege Escalation



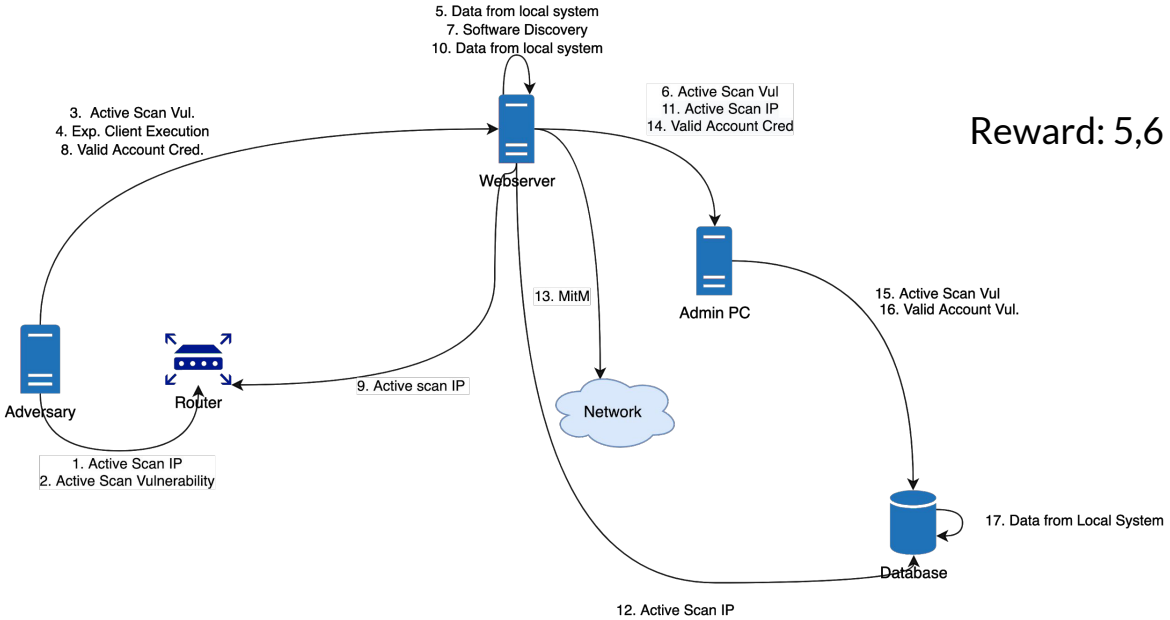
Reward: 2.9

Policy Evaluation Honeypot: Alle Endzustände scheitern



Reward: 0.8

Policy Evaluation - Ziel : Nur Datenbank Auslesen



Reward: 5,6

Policy Evaluation - Beste Policy (kein 0day-Nachteil)

ADVERSARY	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: ROUTER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: ACTIVE_SCAN_VULNERABILITY
	Target: ADMINPC	Action: VALID_ACCOUNTS_VULN
ADMINPC	Target: ADMINPC	Action: DATA_FROM_LOCAL_SYSTEM
	Target: ROUTER	Action: ACTIVE_SCAN_IP_PORT
	Target: DATABASE	Action: ACTIVE_SCAN_IP_PORT
	Target: WEBSERVER	Action: ACTIVE_SCAN_IP_PORT
	Target: ADMINPC	Action: MAN_IN_THE_MIDDLE
	Target: WEBSERVER	Action: ACTIVE_SCAN_VULNERABILITY
	Target: DATABASE	Action: ACTIVE_SCAN_VULNERABILITY
	Target: WEBSERVER	Action: VALID_ACCOUNTS_CRED
WEBSERVER	Target: DATABASE	Action: VALID_ACCOUNTS_CRED
DATABASE	Target: DATABASE	Action: DATA_FROM_LOCAL_SYSTEM

Aktionen: **14**

Reward: **5.9**

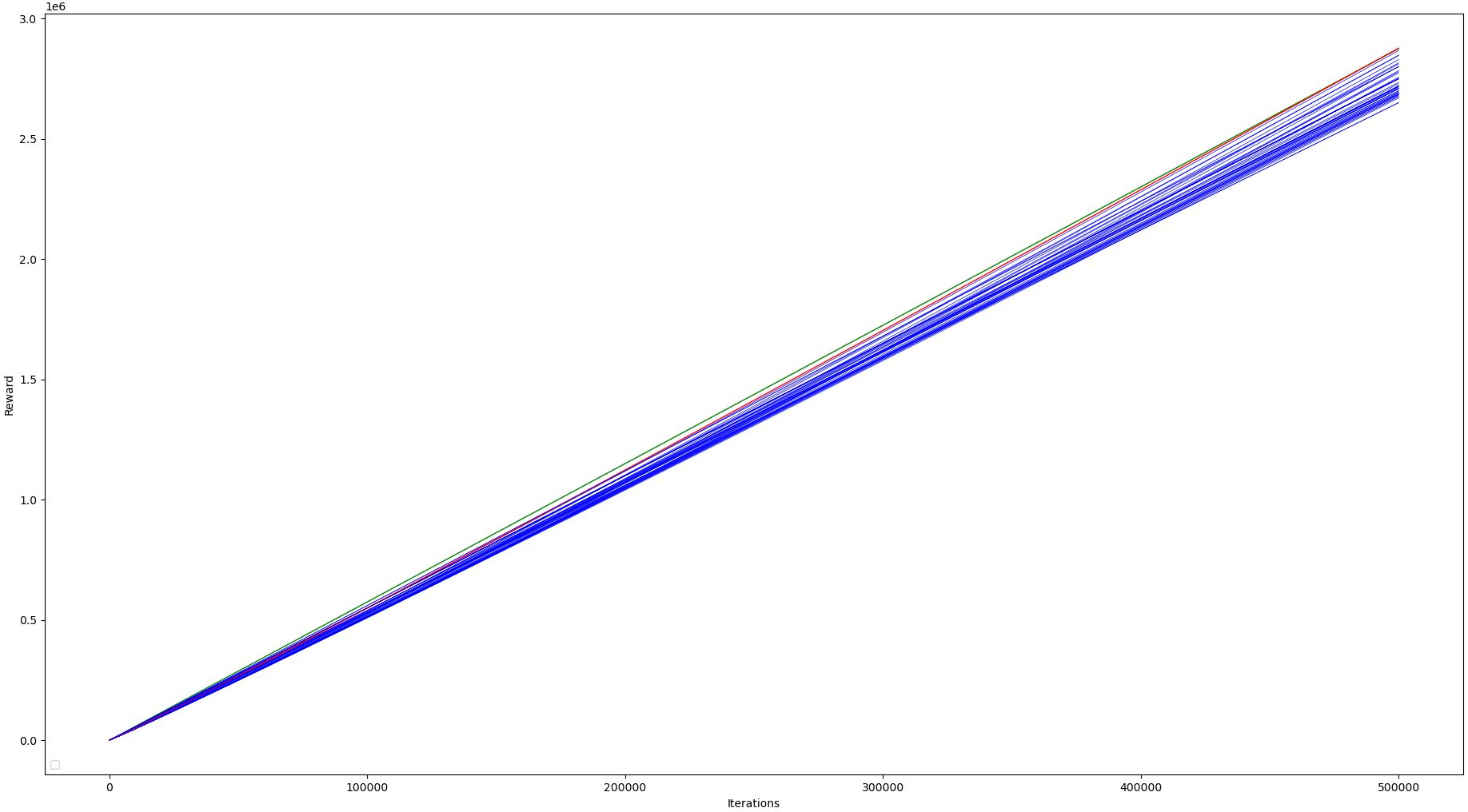
Beste Policy mit Zeroday Nachteil: siehe Folie 14

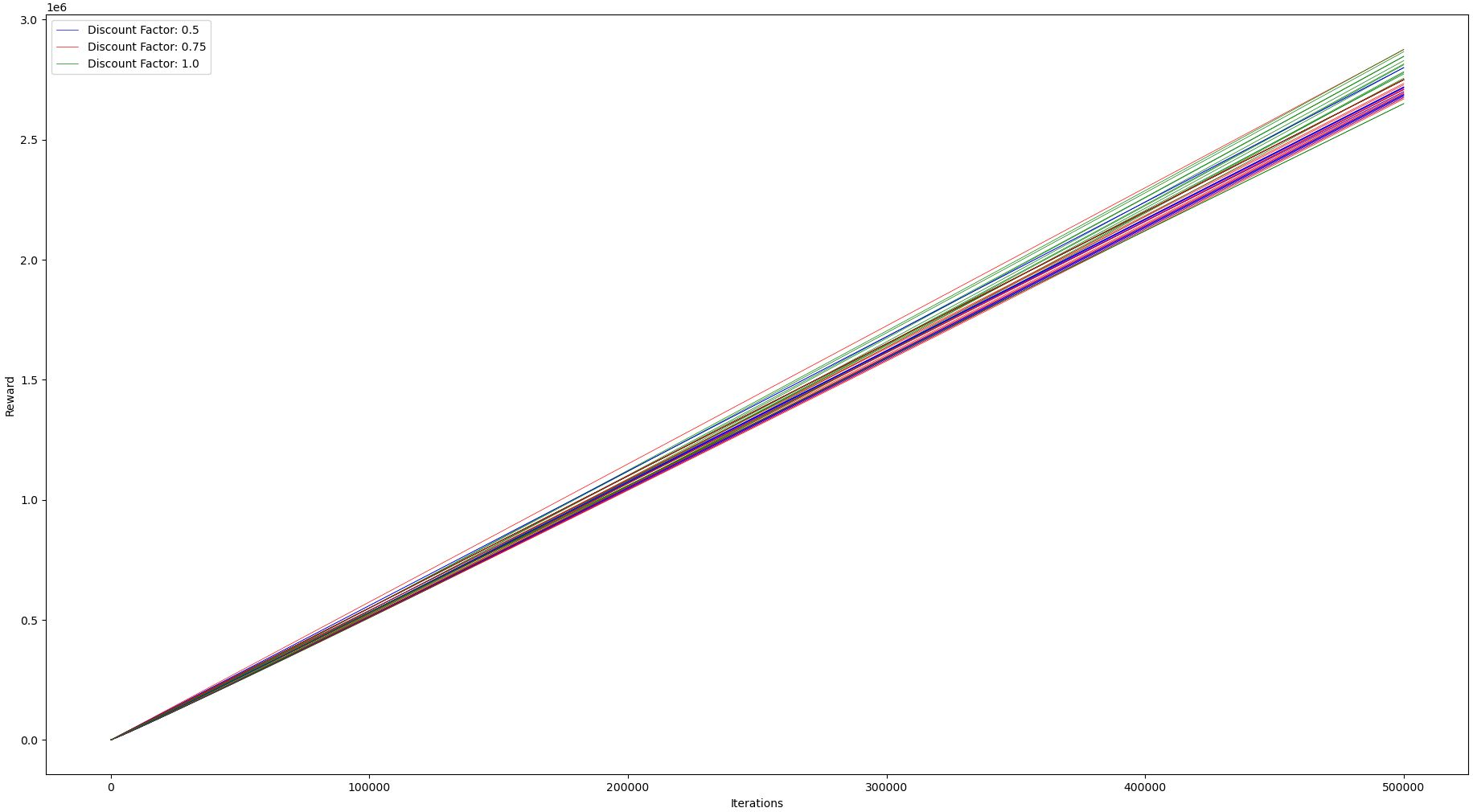
Learning Evaluation

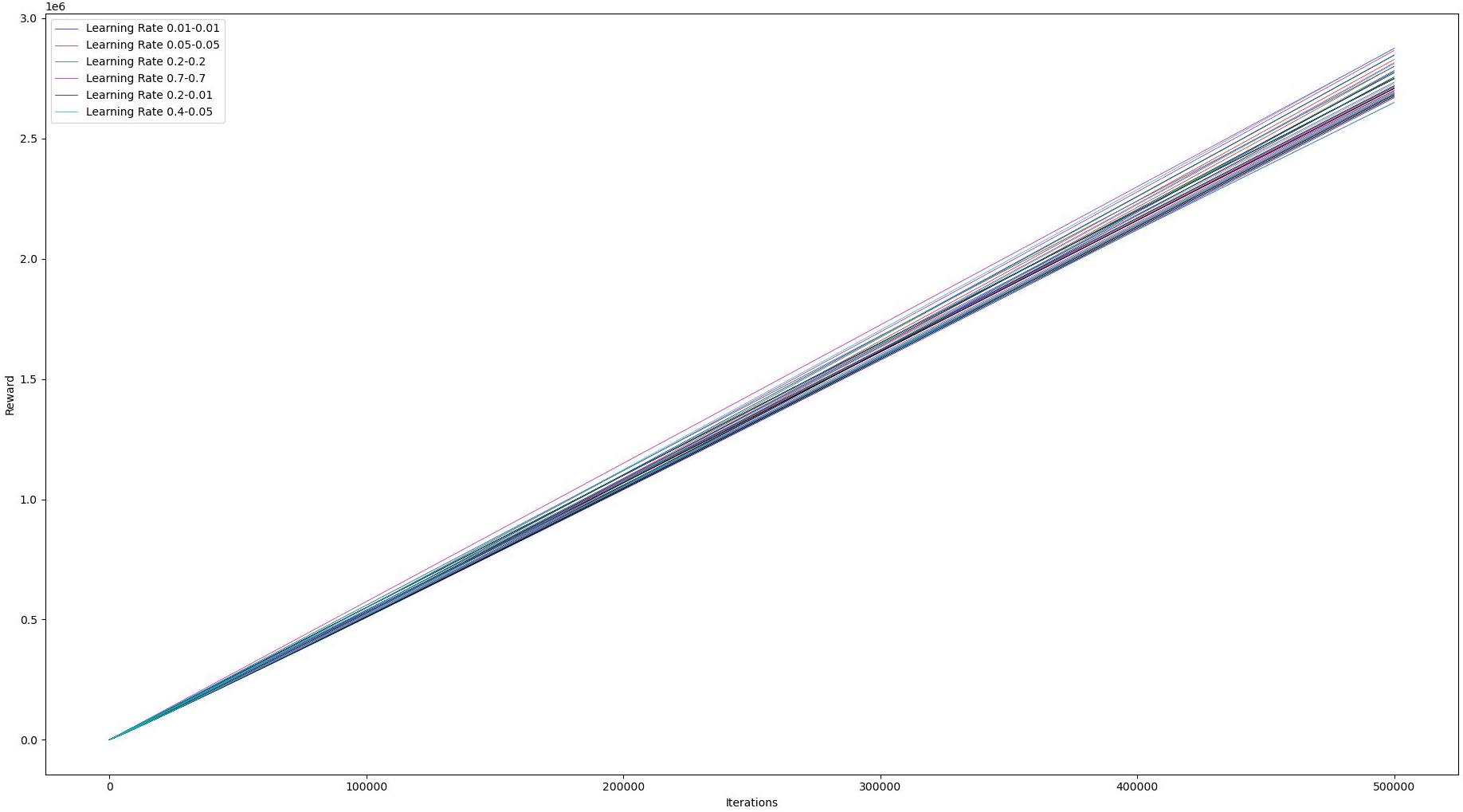
- Vergleich der kumulierten Rewards mit verschiedenen Parametern
  - 54 Parameter Kombinationen
- Test mit “Normaler” ohne 0day und zwischenziel-orientierter Rewardfunktion
- Zwischenziel-Orientierte Rewardfunktion:
  - Zielzustand erhält keinen Reward
  - Aktionen haben die üblichen Kosten
  - Root Zugang erhält einen großen Bonus
  - User Zugang erhält einen mittleren Bonus
  - Daten der Datenbank erhalten erhält einen mittleren Bonus

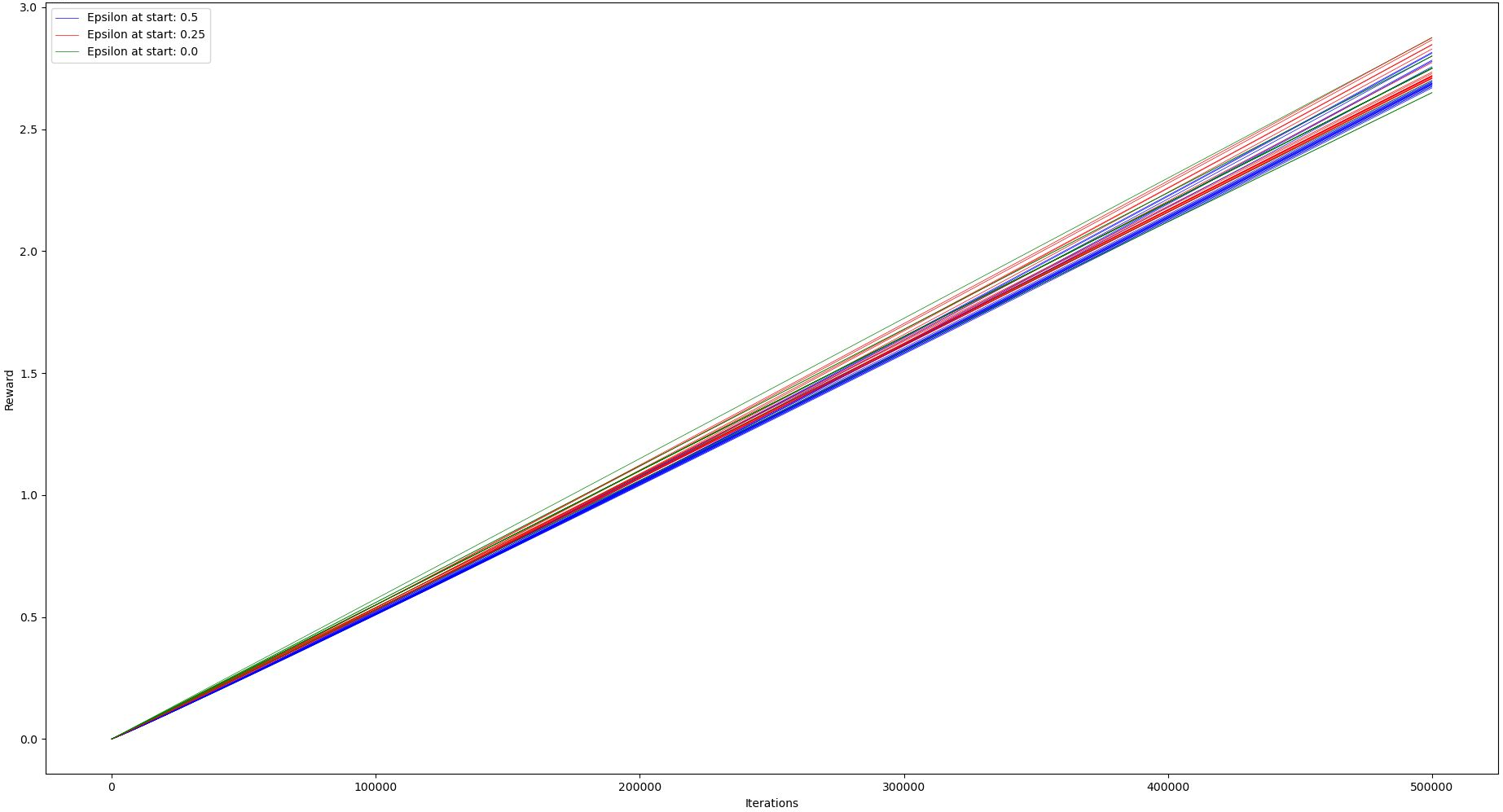
Learning Evaluation - Parameter

- Learning Rate:
  - Statisch:
    - 1%, 5%, 20%, 70%
  - Linear abfallend mit Besuchzahl für jeden Zustand
  - nach 10 Besuchen Minimum erreicht
    - Von 20% nach 1%, Von 40% nach 5 %
- Epsilon (1-Greediness):
  - Abfallend mit Iterationszahl bis auf 0% (100% Greediness)
    - 50%, 25%, 0%
- Discount Factor:
  - Statisch:
    - 50%, 75%, 100%


1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Reward
0
100000
200000
300000
400000
500000
Iterations


1e6
Discount Factor: 0.5
Discount Factor: 0.75
Discount Factor: 1.0
Reward
Iterations
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
100000
200000
300000
400000
500000


1e6
Learning Rate 0.01-0.01
Learning Rate 0.05-0.05
Learning Rate 0.2-0.2
Learning Rate 0.7-0.7
Learning Rate 0.2-0.01
Learning Rate 0.4-0.05
Reward
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
100000
200000
300000
400000
500000
Iterations


1e6
Epsilon at start: 0.5
Epsilon at start: 0.25
Epsilon at start: 0.0
Reward
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
100000
200000
300000
400000
500000
Iterations

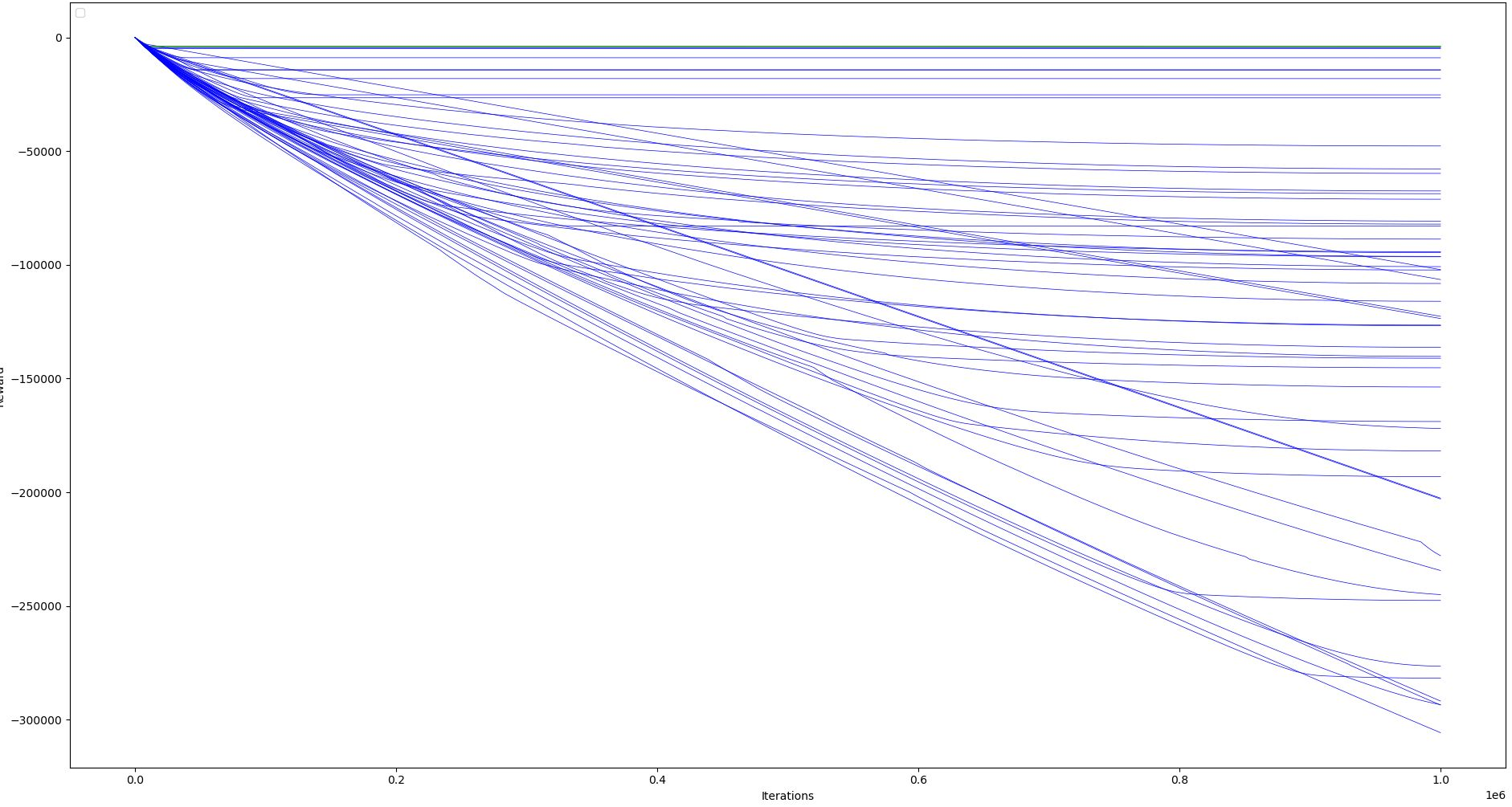
“Normale” Rewardfunktion - Fazit

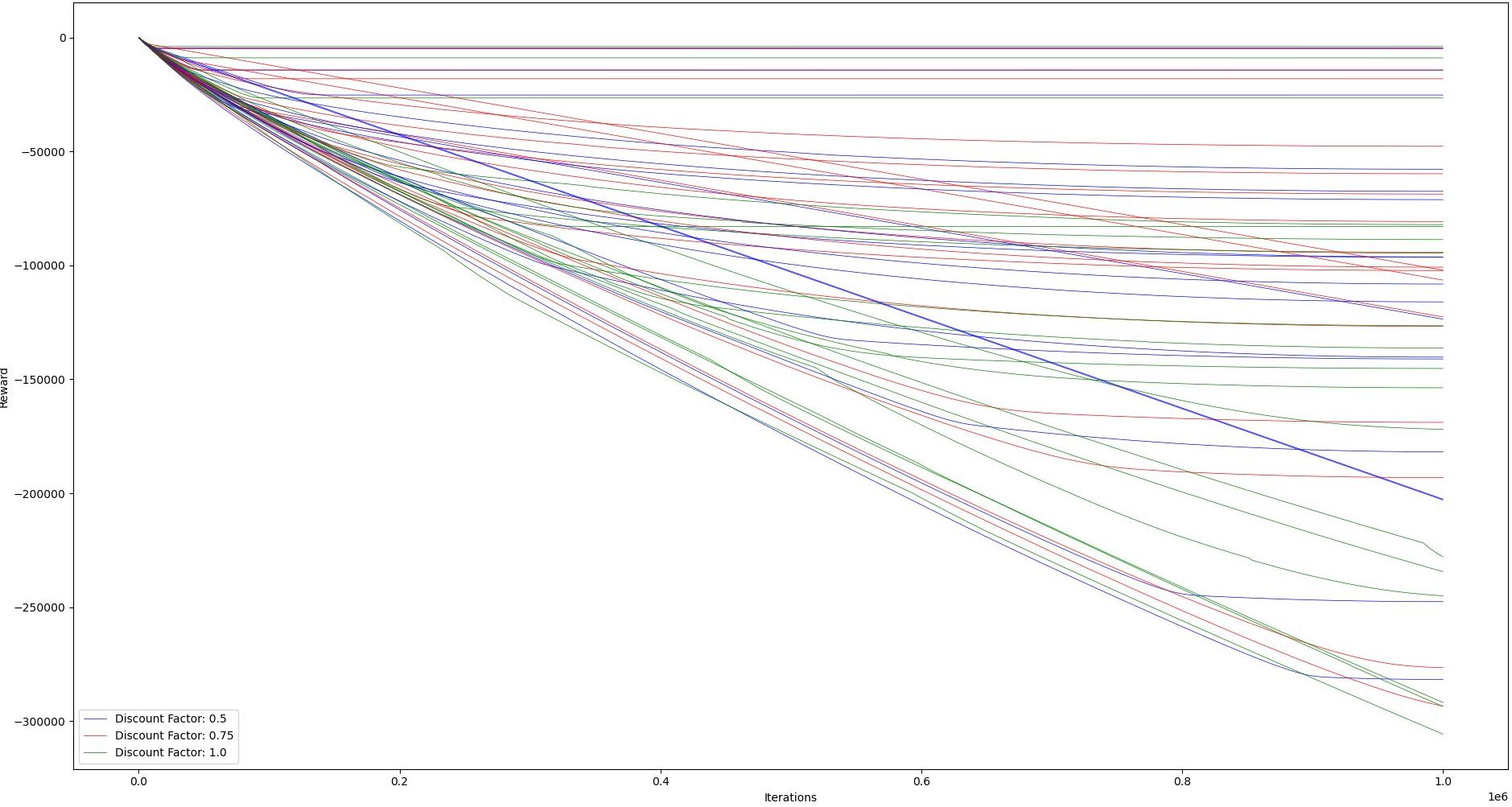
- Hoher Discount Factor führt zu guten Ergebnissen
- Learning Rate hat wenig Einfluss
- Mittlere Werte für Epsilon führen zum schnellen Finden der besten Policy

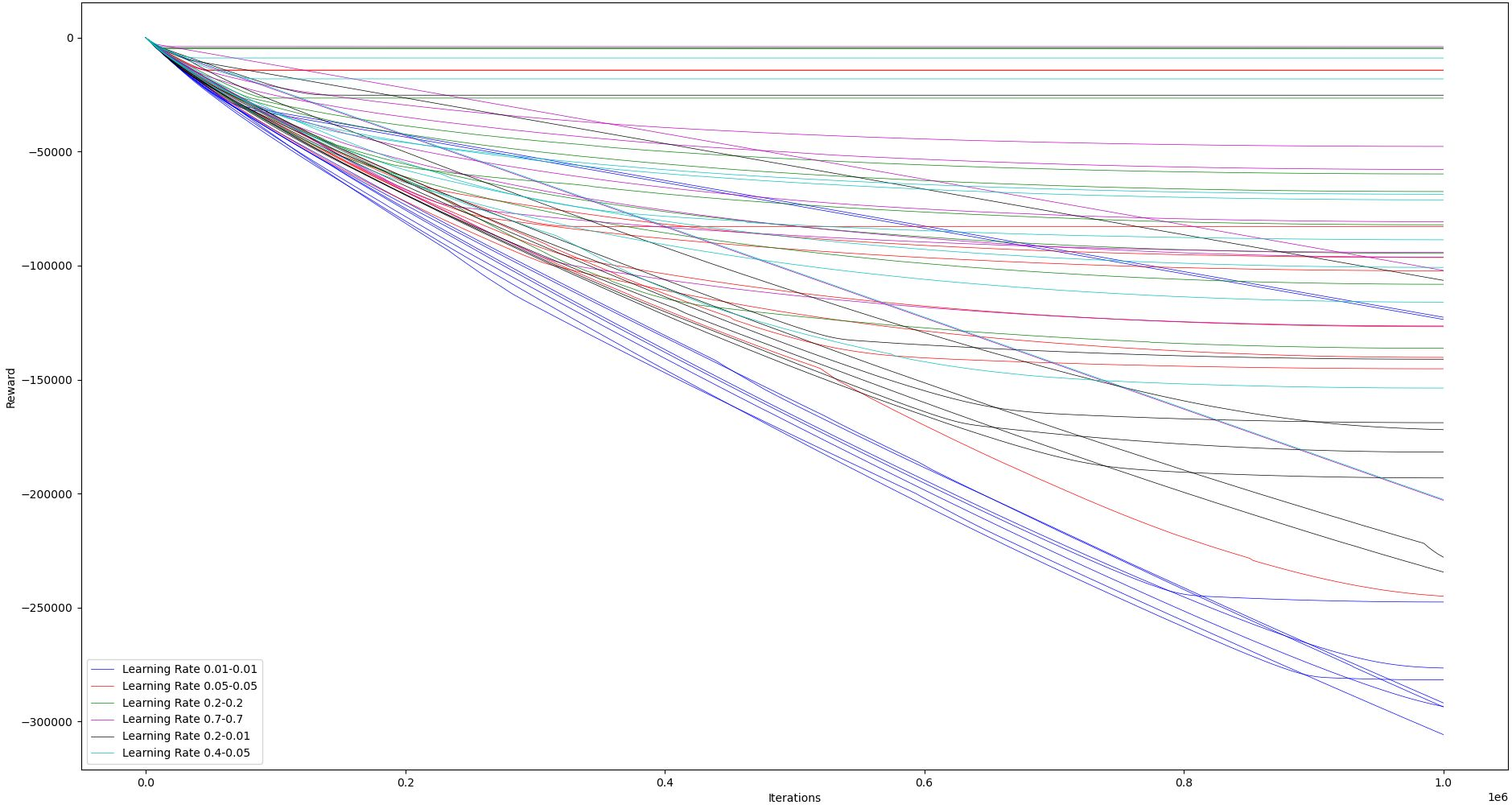
- Parameter der dominanten Kurve:
  - Learning Rate von 40%, abfallend auf 5%
  - Epsilon von 25%, abfallend auf 0%
  - Discount Factor von 100%

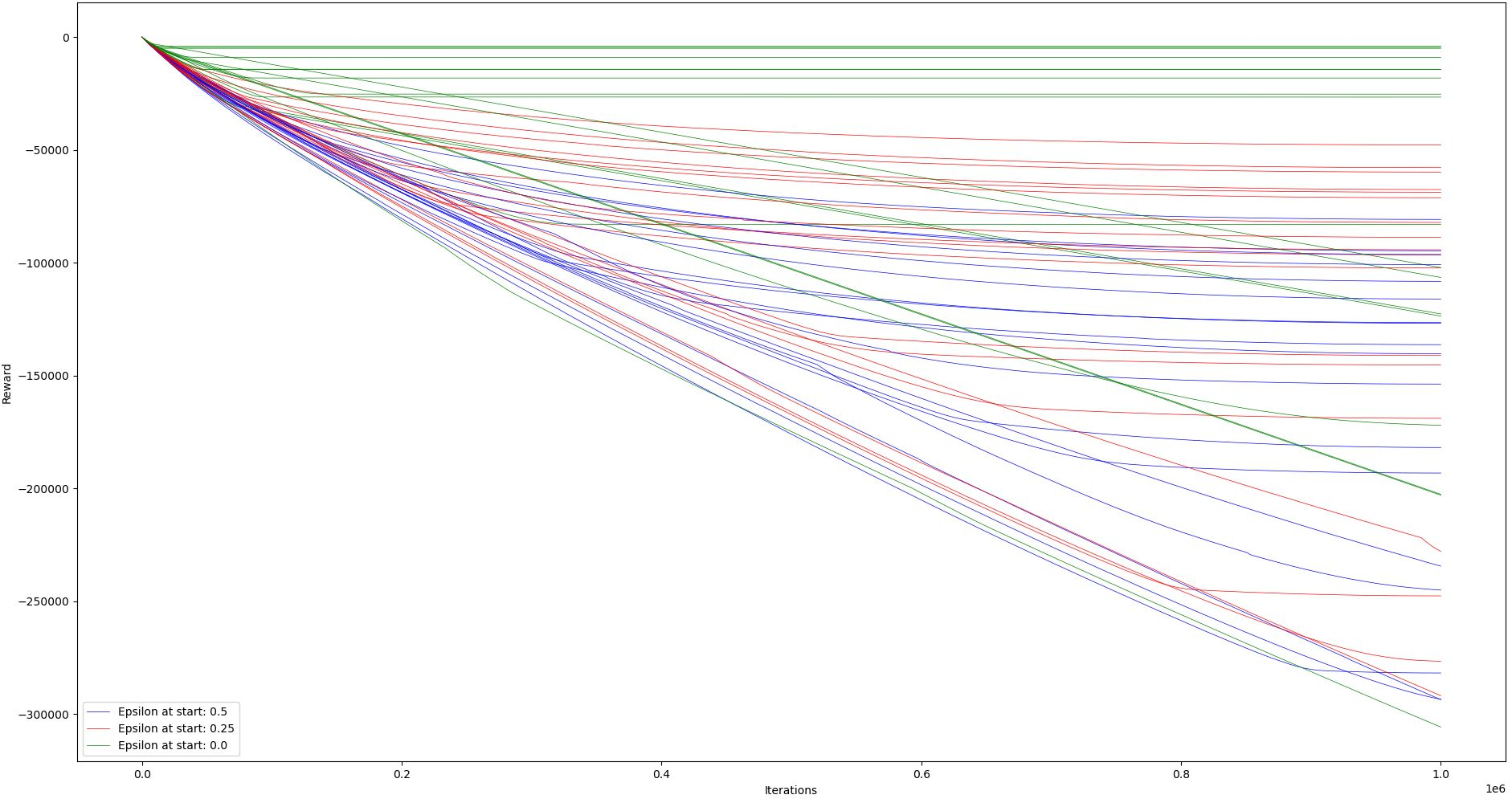
Learning Evaluation

- Zwischenziel-Orientierte Rewardfunktion:
  - Zielzustand erhält keinen Reward
  - Aktionen haben die üblichen Kosten
  - Root Zugang erhält einen großen Bonus
  - User Zugang erhält einen mittleren Bonus
  - Daten der Datenbank erhalten erhält einen mittleren Bonus


0
−50000
−100000
−150000
−200000
−250000
−300000
reward
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Iterations


0
−50000
−100000
−150000
−200000
−250000
−300000
Reward
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Iterations
Discount Factor: 0.5
Discount Factor: 0.75
Discount Factor: 1.0


0
−50000
−100000
−150000
−200000
−250000
−300000
Reward
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Iterations
Learning Rate 0.01-0.01
Learning Rate 0.05-0.05
Learning Rate 0.2-0.2
Learning Rate 0.7-0.7
Learning Rate 0.2-0.01
Learning Rate 0.4-0.05


0
−50000
−100000
−150000
−200000
−250000
−300000
Reward
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Iterations
Epsilon at start: 0.5
Epsilon at start: 0.25
Epsilon at start: 0.0

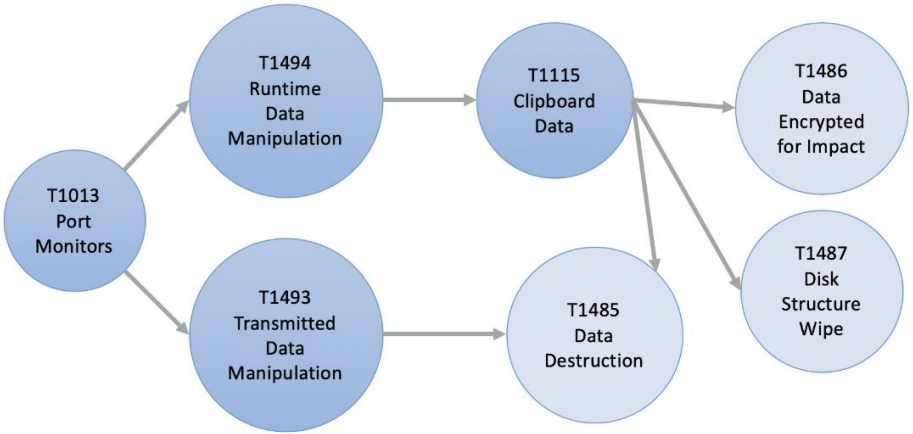
Fazit

- Skalierbarkeit (MDP) problematisch
- Learning Performance dank spezifischer Utility Funktion sehr schnell, aber nicht skalierbar/ganzheitlich/korrekt
- Szenario (Komplexität & Abstraktionsgrad) -> für Anwendbarkeit: **Low Level Beschreibungen**
- benötigen ganzheitliche Abhängigkeiten und Ausformulierung von **Vor- und Nachbedingungen** ATT&CK Techniken (Ansatz z.b. *Al-Shaer et al. [1]*)
- Verknüpfung von ATT&CK Techniken und Angreifer-Vorgehen anhand von **Datensets** erwünscht (Data-driven Prediction siehe [3]), zusätzlich Validierung & Evaluation bzw. Anwendbarkeit
- Erkennen des Angreifer Ziels: Forschungsgebiet Intention Recognition

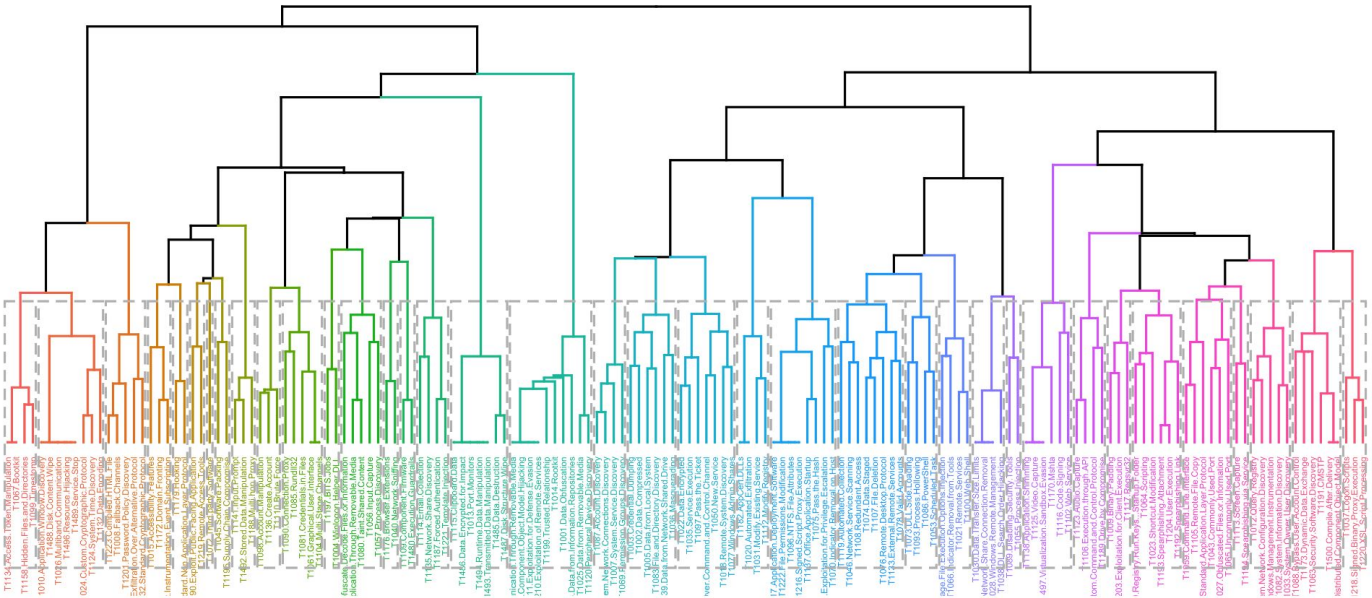
Stand der Technik [1]

[1] Al-Shaer et al.

- Berechnen der Assoziationen von ATT&CK Techniken (TTPs) mit hierarchischem Clustering
- Berechnete Cluster beinhalten sequentielle, disjunktive und konjunktive Beziehungen (benötigt immer noch Expertenwissen)
- Mitre Datensatz Probleme: Vollständigkeit, Richtigkeit, Bias
- Zwar auch Beziehungen zw. einzelnen Clustern erkannt, aber keine ganzheitlichen Abhängigkeiten



Al-Shaer et al. [1] - ATT&CK Assoziationen

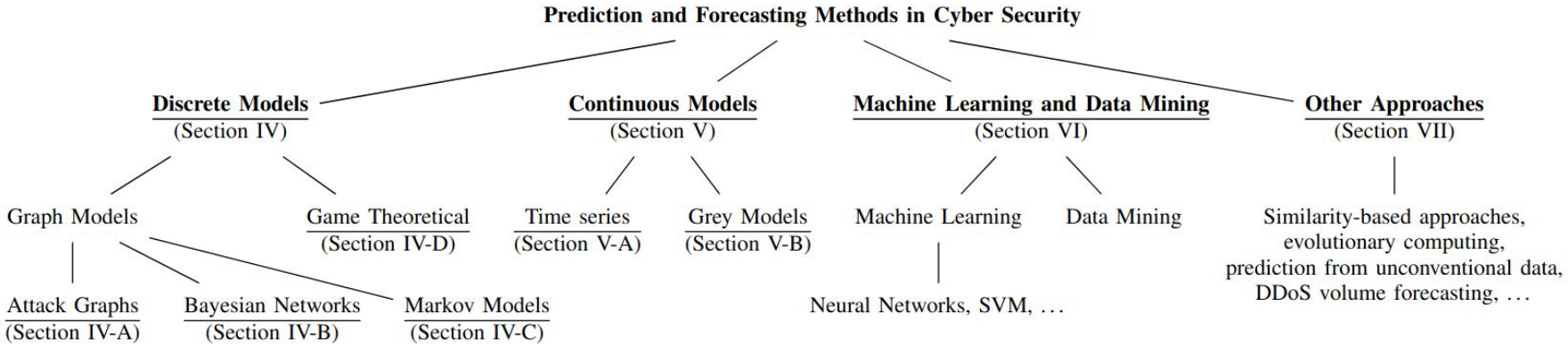


Stand der Technik [2] - Survey

Hus´ak et al.

- Für Vorhersage ist die grobe Angriffssequenz bekannt, aber wir brauchen formale Beschreibung
- Neben Prediction/Projection ist insbesondere das Gebiet Forecasting interessant (Network Situational Awareness)
- Gängige Ansätze: siehe nächste Folie
- Modell-basiert v.a. für Angriff Vorhersage, Continuous für Situational Awareness Bewertung
- Aktuell beliebt: **ML & Data Mining**, im Kommen: Deep Learning, **Big Data**, Collaborative IDS
- Wie wirken sich neue Paradigmen aus: IoT, SDN, ...

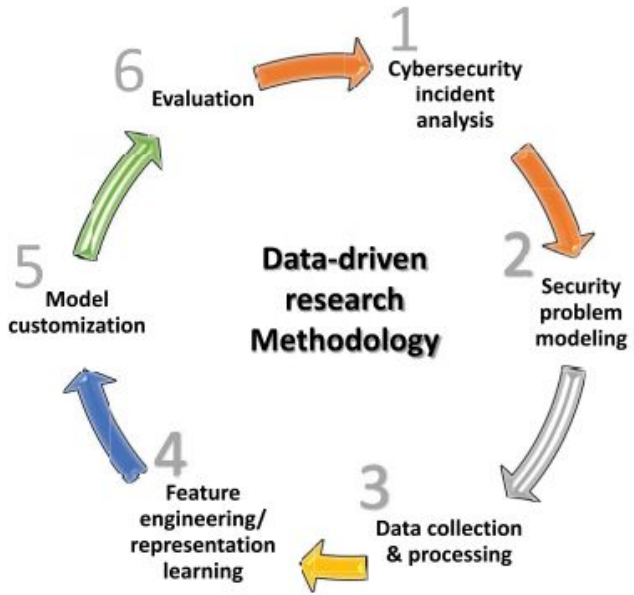
Methoden für Prediction, Forecasting [2]



Stand der Technik [3] - Incident Prediction und Daten

Sun et al.

- Daten als essentieller Bestandteil, beinflussen Modelle und Modellbildung
- Klassifikation von Datensets: Organisations-Reports, Netzwerkdaten, Soziale Medien, Synthetische Daten, Webcrawler …
- Anders als bei uns meist **spezifische Incidents** (z.b. Hidden Sensitive Operations finden)
- Erkenntnisse: wertvolle Daten meist schwer zu finden, daher verschiedene Quellen bzw. Blickwinkel korrelieren/aggregieren
- Datenset Qualität (Bias, Verlässlichkeit/Genauigkeit, Vollständigkeit)
- Representation Learning als erfolgversprechende Alternative zu Feature Engineering



Literatur

[1] R. Al-Shaer, J. M. Spring, and E. Christou. Learning the associations of mitre att&ck adversarial techniques. In 2020 IEEE Conference on Communications and Network Security (CNS), pages 1–9, June 2020.

[2] M. Hus´ak, J. Kom´arkov´a, E. Bou-Harb, and P. Cˇeleda. Survey of Attack Projection, Prediction, and Forecasting in Cyber Security. IEEE Communications Surveys & Tutorials, 21(1):640–660, 2019.

[3] N. Sun, J. Zhang, P. Rimba, S. Gao, L. Y. Zhang, and Y. Xiang. Data-driven cybersecurity incident prediction: A survey. IEEE Communications Surveys & Tutorials, 21(2):1744–1772, Second quarter 2019.