Självlärande System för Cyberförsvar

CDIS Besök av försvarsdepartementet

Kim Hammar

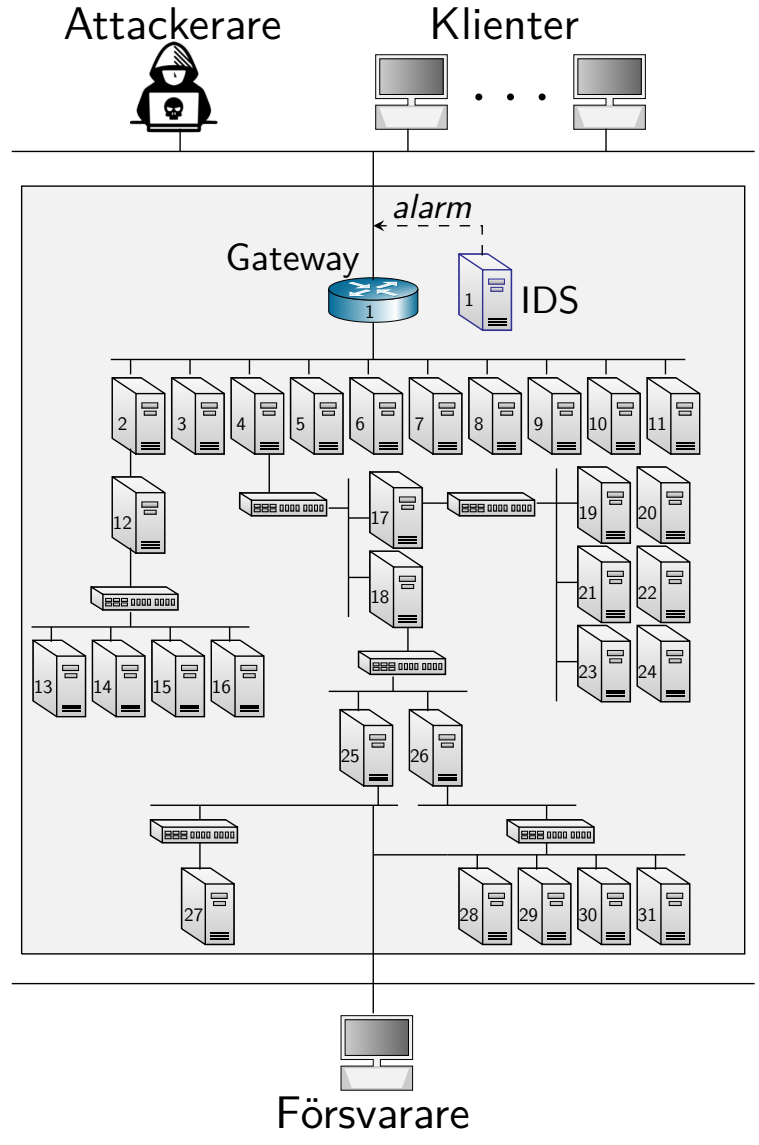
kimham@kth.se

CDIS, Centrum för cyberförsvar och informationssäkerhet
NSE, Avdelningen för nätverk och systemteknik
KTH Kungliga Tekniska Högskolan

21 Feb, 2023

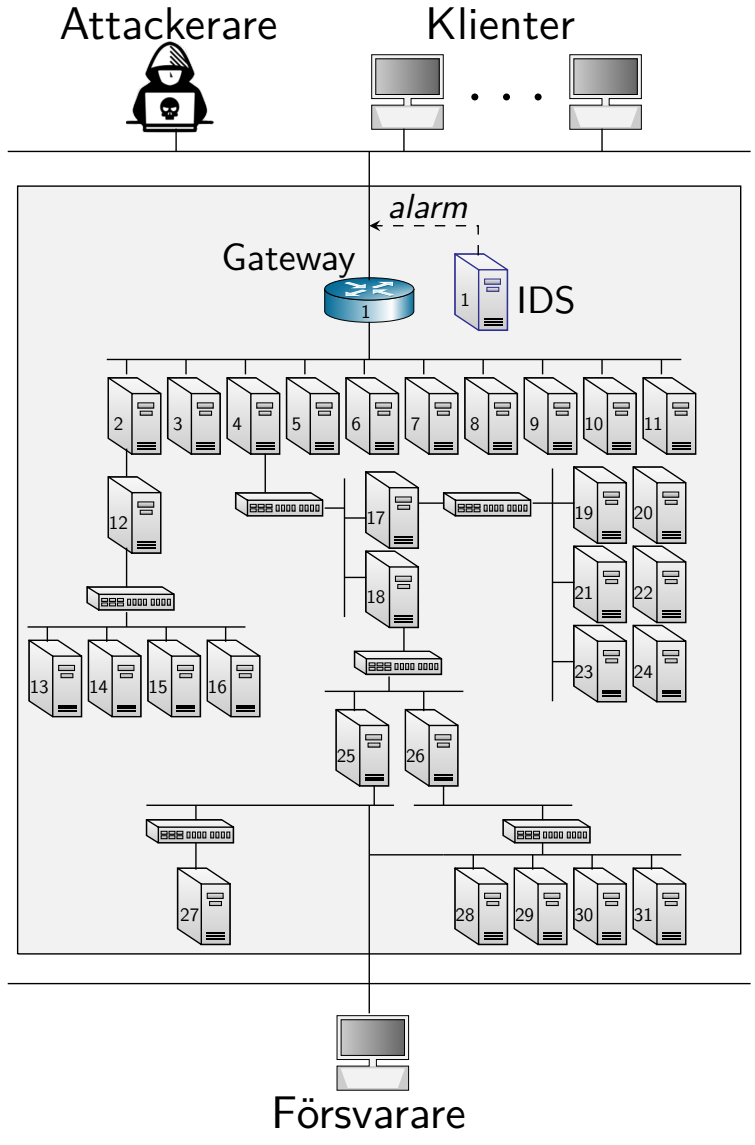
Utmaning: Automatiserade och föränderliga attackmetoder

- **Utmaningar**:
  - Attackmetoder är i en konstant förändring och utveckling
  - Komplicerade IT-infrastrukturer

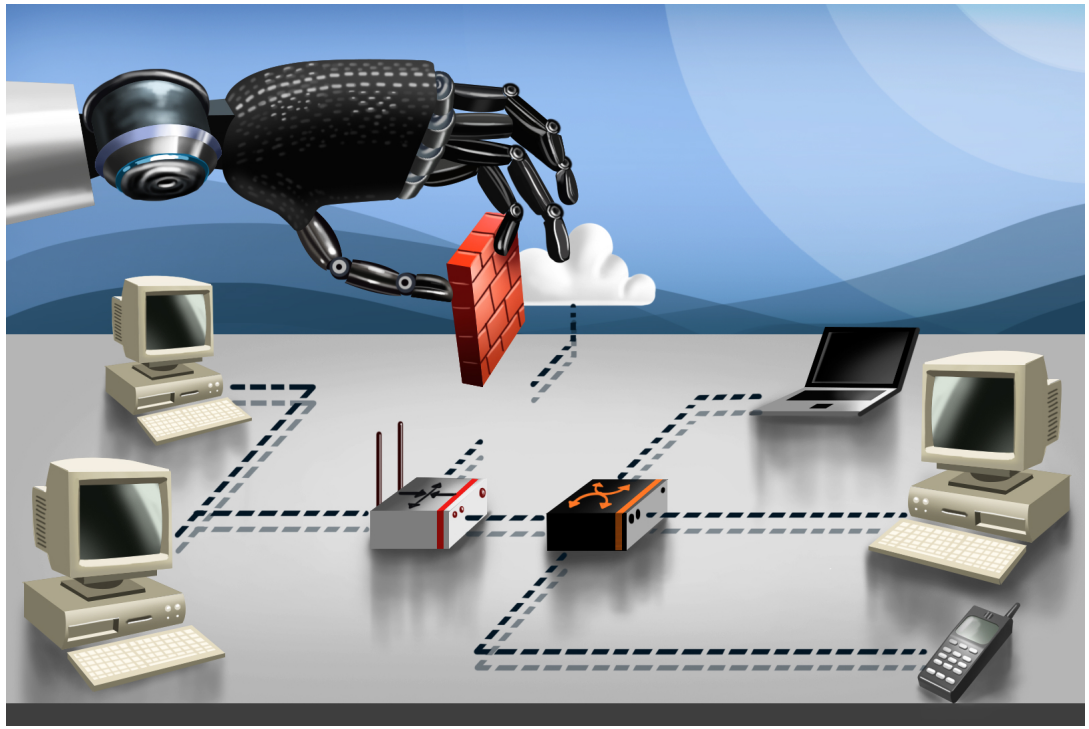


Forskningsmål: Automatiserad säkerhet och inlärning

- **Utmaningar**:
  - Attackmetoder är i en konstant förändring och utveckling
  - Komplicerade IT-infrastrukturer
- **Forskningsmål**:
  - Automatisera säkerhetsfunktioner
  - Anpassa system till föränderliga attackmetoder



Automatiserad Säkerhet: Nuvarande Forskningslandskap



Nivåer av säkerhetsautomatisering

Ingen automatisering.
Manuell detektering.
Manuell prevention.
Inga alarm.
Ingen automatiserad attack mitigering.
Brist på verktyg.

80-talet

Operatörassistans.
Manuell detektering.
Manuell prevention.
Granskningsloggar.
Säkerhetsverktyg.

90-talet

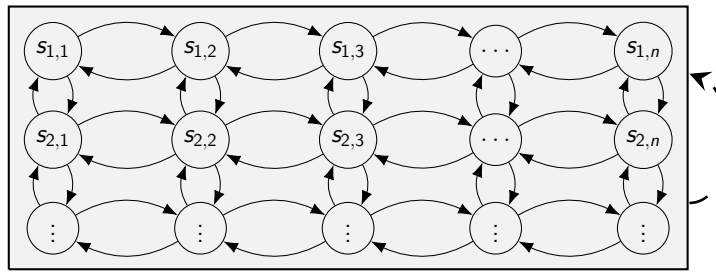
Partiell automatisering.
System har automatiserade funktioner för detektering/ prevention men kräver manuell uppdatering och konfiguration.
Intrångsdetekteringssystem.
Intrångspreventeringssystem.

00-talet-Nu

Hög automatisering.
Systemet uppdaterar sig självt automatiskt.
Automatiserad attackdetektering.
Automatiserad attackmitigering.

Forskning

Vår metod för att automatiskt beräkna säkerhetsstrategier

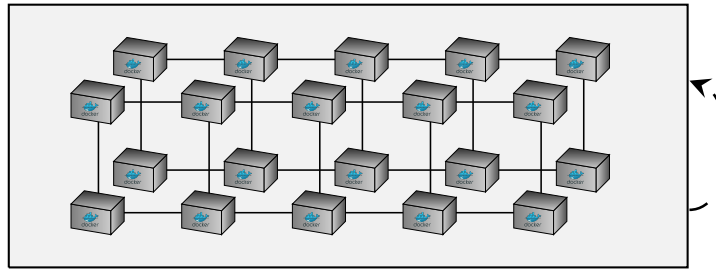


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

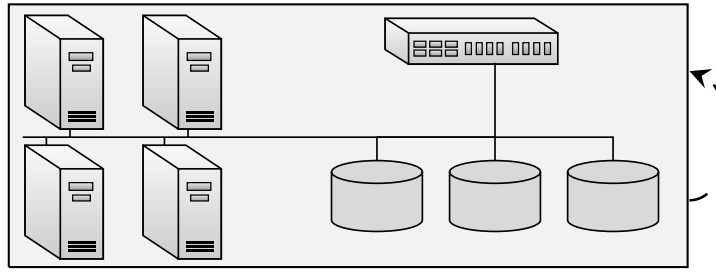


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

Selektiv replikering

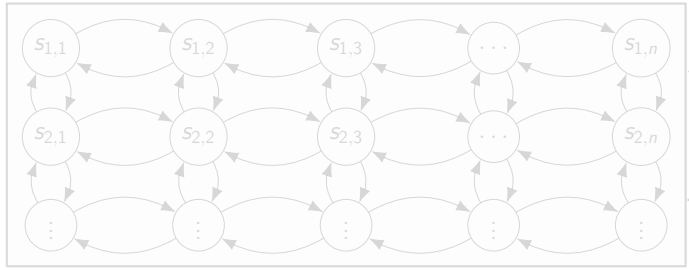


IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

SIMULERINGSSYSTEM

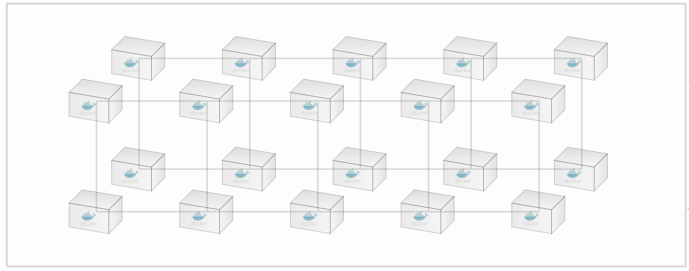


Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

EMULERINGSSYSTEM

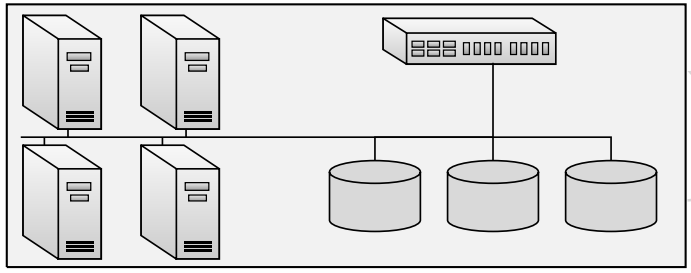


Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

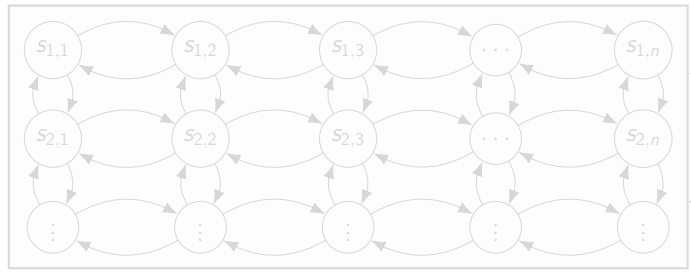
Selektiv replikering

IT INFRASTRUKTUR



Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

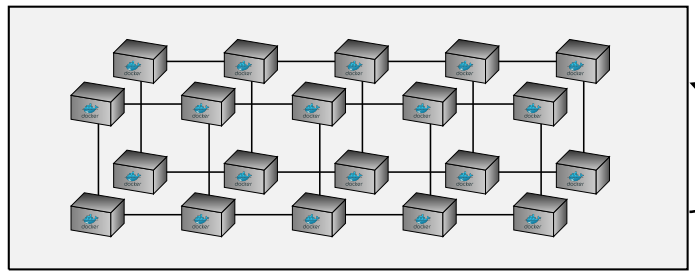


Simuleringssystem

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



Emuleringssystem

Utvärdering & skattning av modeller

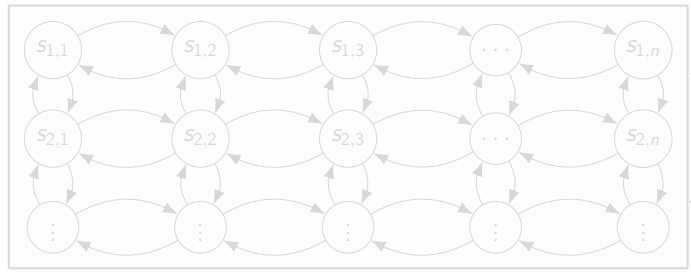
Implementation av säkerhetsstrategi π

Selektiv replikering

IT infrastruktur

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

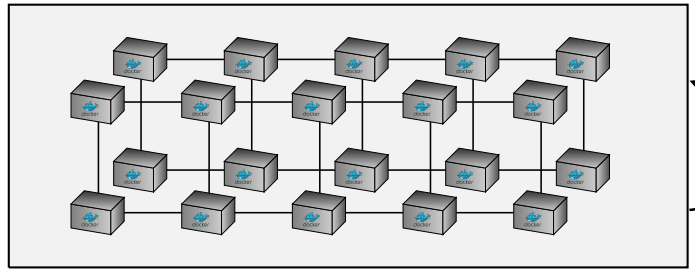


Simuleringssystem

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



Emuleringssystem

Utvärdering & skattning av modeller

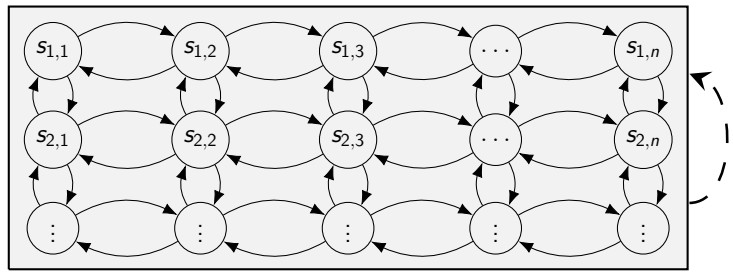
Implementation av säkerhetsstrategi π

Selektiv replikering

IT infrastruktur

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

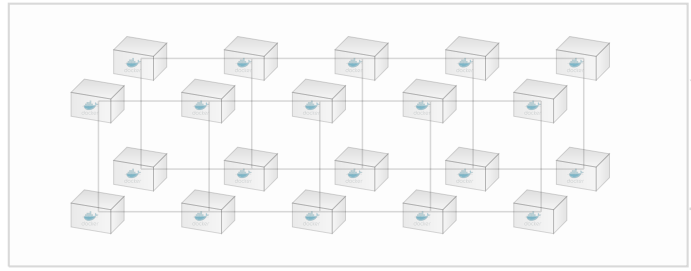


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



EMULERINGSSYSTEM

Utvärdering & skattning av modeller

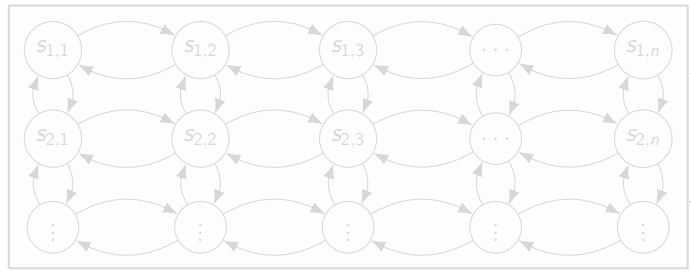
Implementation av säkerhetsstrategi π

Selektiv replikering

IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

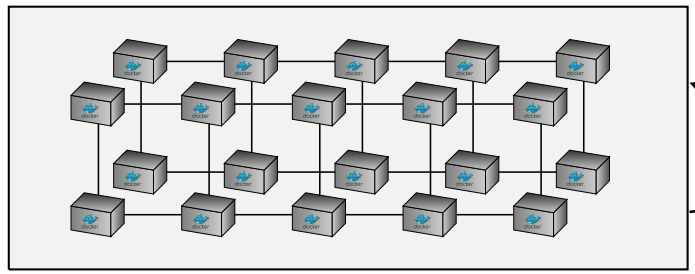


Simuleringssystem

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



Emuleringssystem

Utvärdering & skattning av modeller

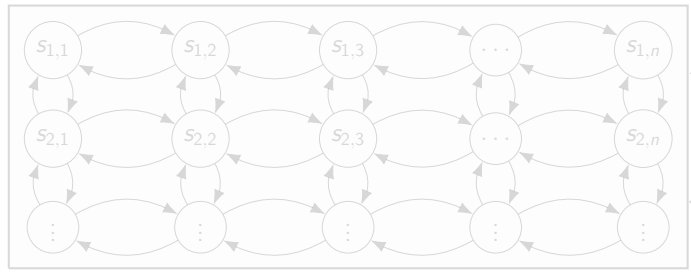
Implementation av säkerhetsstrategi π

Selektiv replikering

IT infrastruktur

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

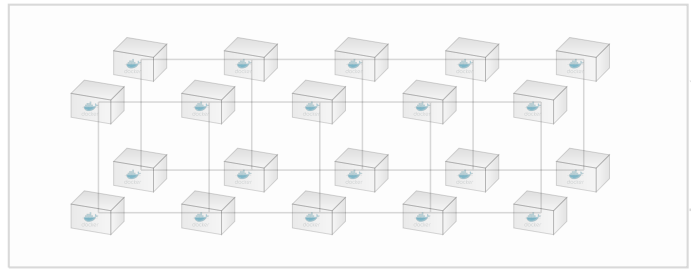


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

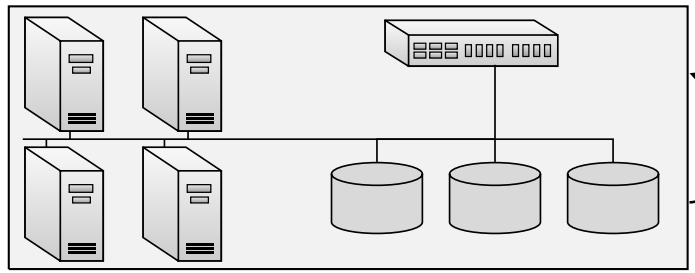


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

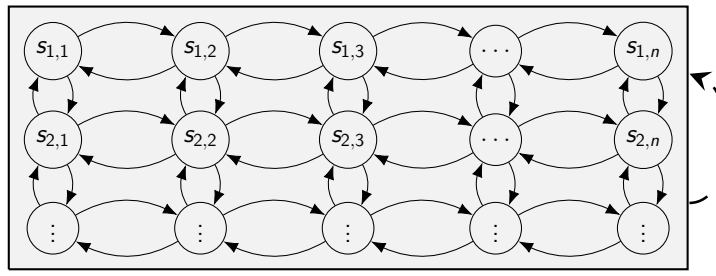
Selektiv replikering



IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

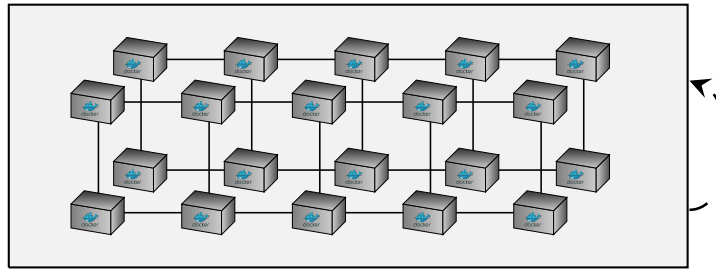


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

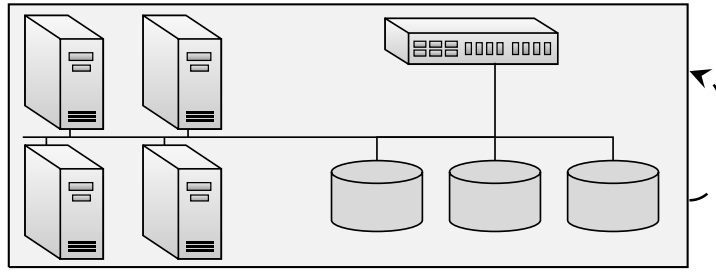


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

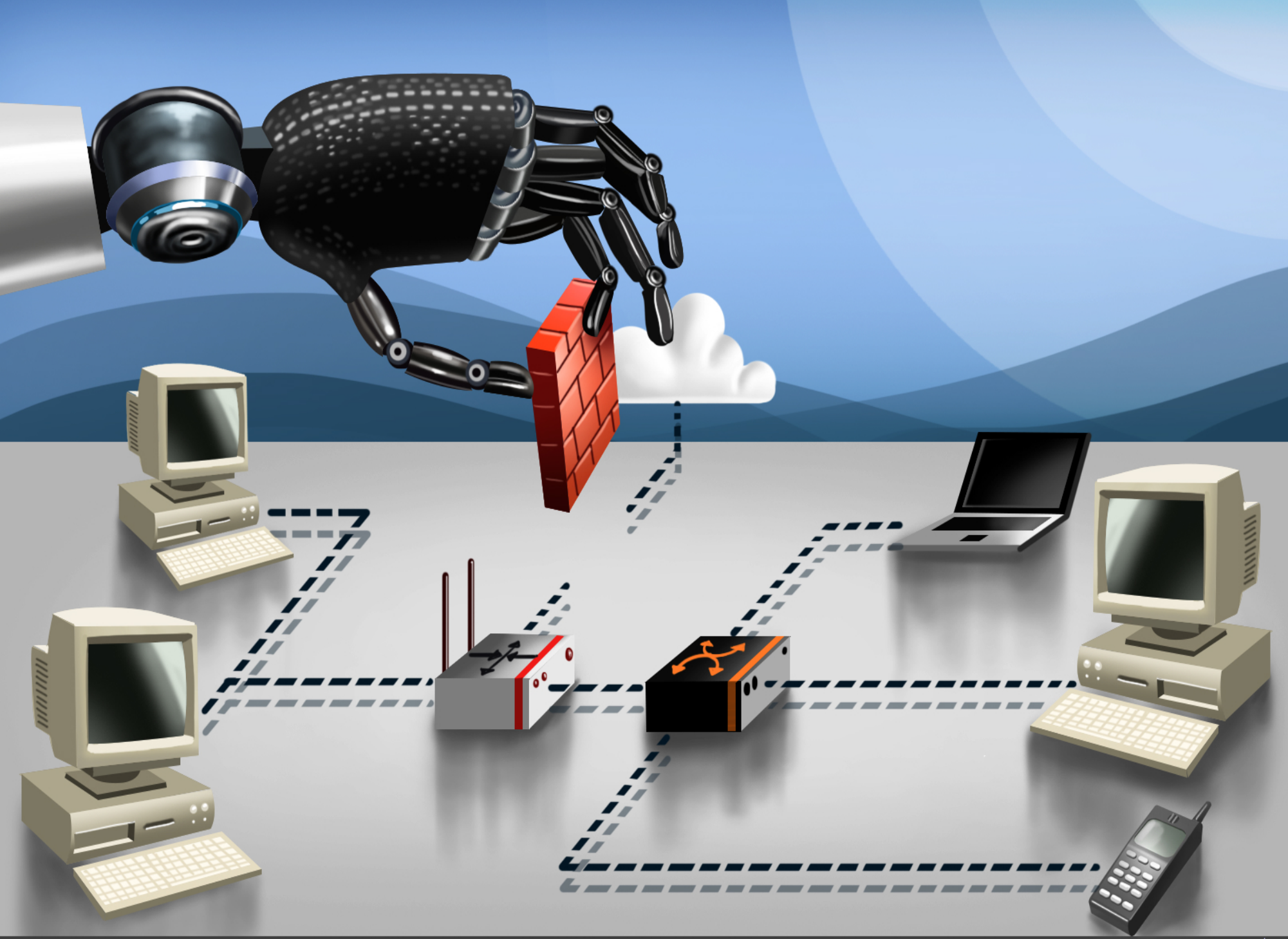
Implementation av säkerhetsstrategi π

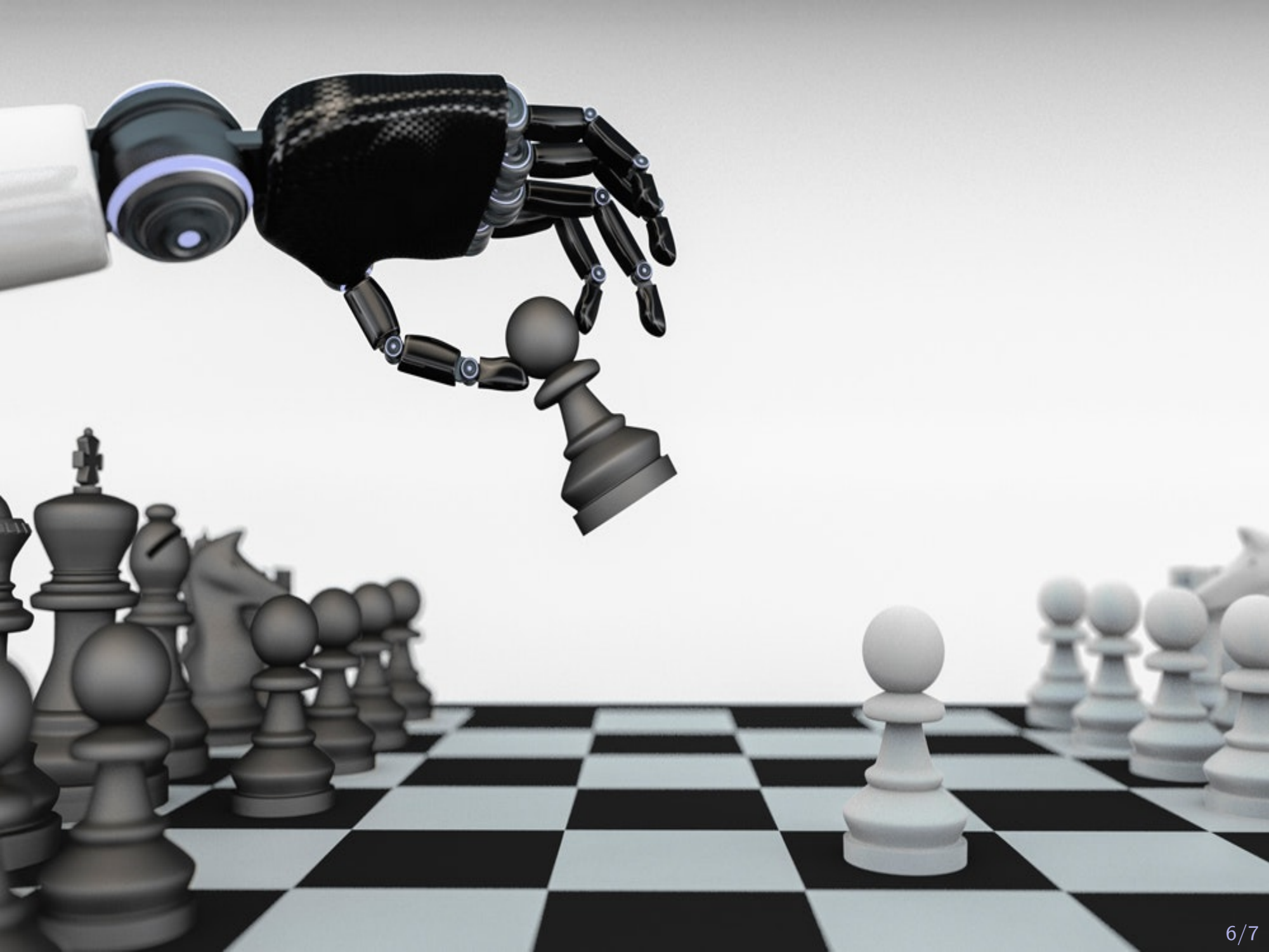
Selektiv replikering



IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system





Referenser

- *Learning Near-Optimal Intrusion Responses Against Dynamic Attackers*[1]
- *Finding Effective Security Strategies through Reinforcement Learning and Self-Play*[2]
- *Learning Intrusion Prevention Policies through Optimal Stopping*[3]
- *A System for Interactive Examination of Learned Security Policies*[4]
- *Intrusion Prevention Through Optimal Stopping*[5]
- *Learning Security Strategies through Game Play and Optimal Stopping*[6]

---

[1] Kim Hammar and Rolf Stadler. *Learning Near-Optimal Intrusion Responses Against Dynamic Attackers*. 2023. DOI: 10.48550/ARXIV.2301.06085. URL: https://arxiv.org/abs/2301.06085.

[2] Kim Hammar and Rolf Stadler. "Finding Effective Security Strategies through Reinforcement Learning and Self-Play". In: *International Conference on Network and Service Management (CNSM 2020)*. Izmir, Turkey, 2020.

[3] Kim Hammar and Rolf Stadler. "Learning Intrusion Prevention Policies through Optimal Stopping". In: *International Conference on Network and Service Management (CNSM 2021)*. http://dl.ifip.org/db/conf/cnsm/cnsm2021/1570732932.pdf. Izmir, Turkey, 2021.

[4] Kim Hammar and Rolf Stadler. "A System for Interactive Examination of Learned Security Policies". In: *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*. 2022, pp. 1–3. DOI: 10.1109/NOMS54207.2022.9789707.

[5] Kim Hammar and Rolf Stadler. "Intrusion Prevention Through Optimal Stopping". In: *IEEE Transactions on Network and Service Management* 19.3 (2022), pp. 2333–2348. DOI: 10.1109/TNSM.2022.3176781.

[6] Kim Hammar and Rolf Stadler. "Learning Security Strategies through Game Play and Optimal Stopping". In: *Proceedings of the ML4Cyber workshop, ICML 2022, Baltimore, USA, July 17-23, 2022*. PMLR, 2022.