Självlärande system för cybersäkerhet

CDIS Besök av riksdagens försvarsutskott (S)

Kim Hammar, Doktorand

Handledare: Prof. Rolf Stadler & Prof. Pontus Johnson

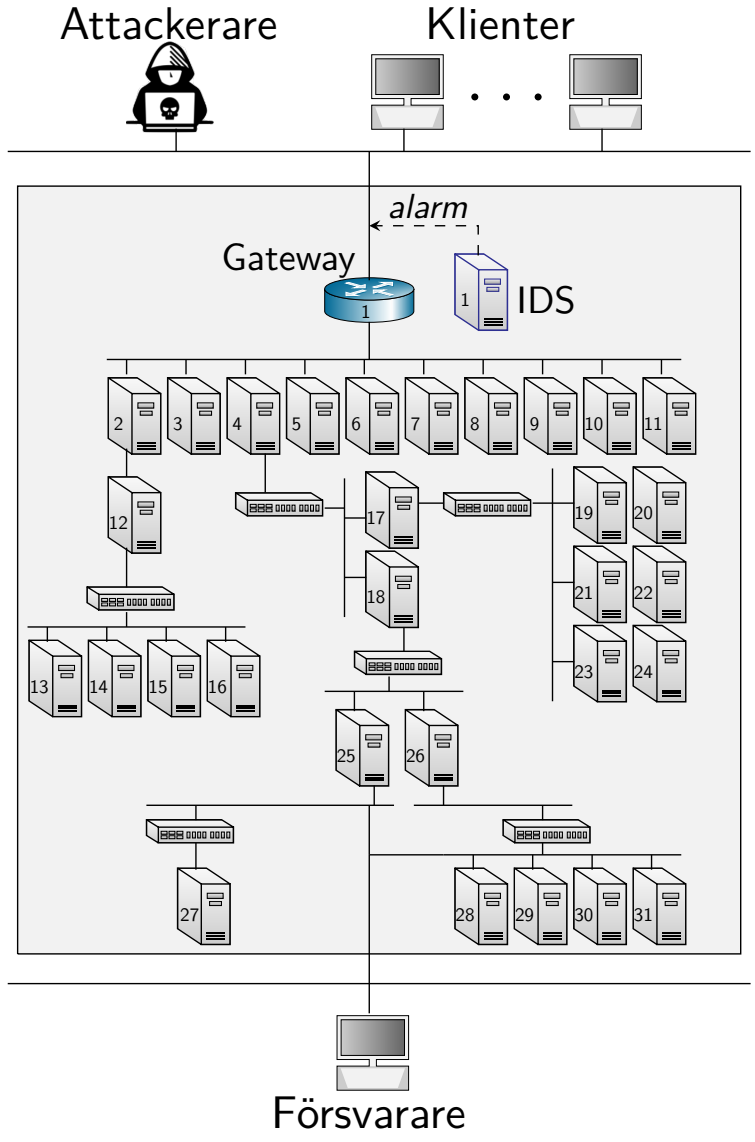
kimham@kth.se

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

20 Okt, 2021

Utmaning: Automatiserade och föränderliga attackmetoder

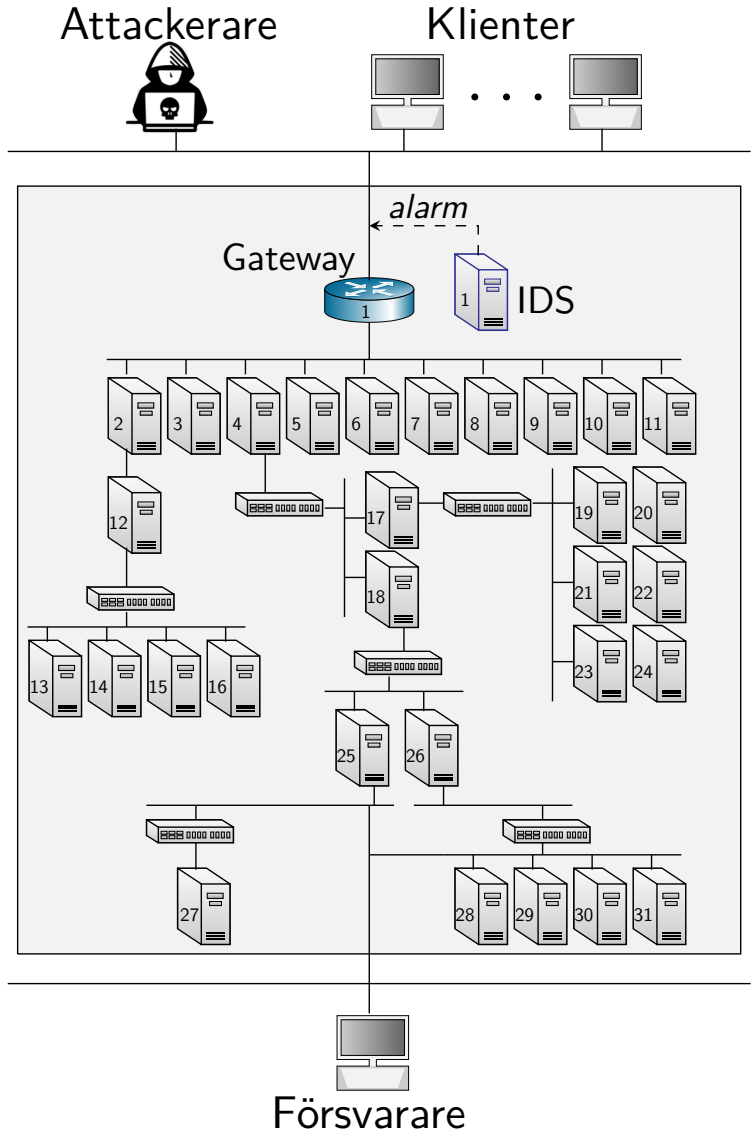
- **Utmaningar**:
  - Attackmetoder är i en konstant förändring och utveckling
  - Komplicerade IT-infrastrukturer



Forskningsmål: Automatiserad säkerhet och inlärning

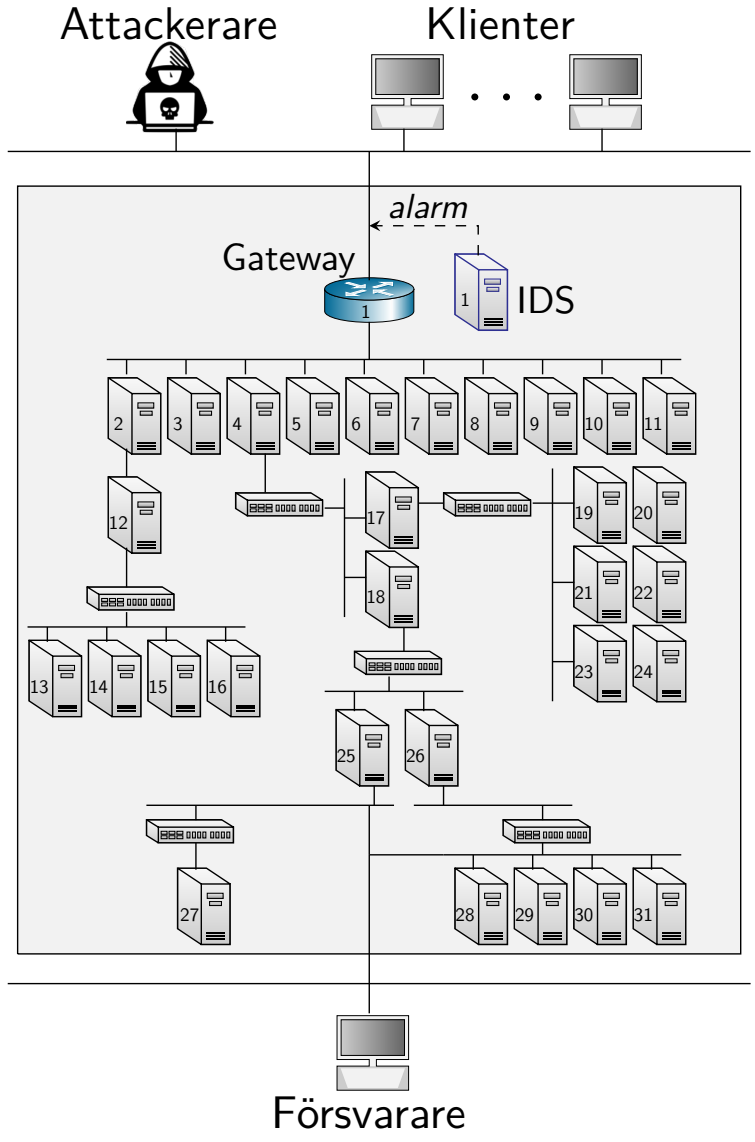
- **Utmaningar**:
  - Attackmetoder är i en konstant förändring och utveckling
  - Komplicerade IT-infrastrukturer

- **Forskningsmål**:
  - Automatisera säkerhetsfunktioner
  - Anpassa system till föränderliga attackmetoder

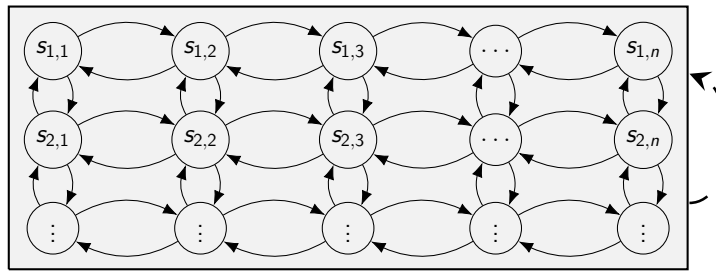


Metod: formella modeller & förstärkningsinlärning

- **Utmaningar**:
  - Attackmetoder är i en konstant förändring och utveckling
  - Komplicerade IT-infrastrukturer
- **Forskningsmål**:
  - Automatisera säkerhetsfunktioner
  - Anpassa system till föränderliga attackmetoder
- **Forskningsmetod**:
  - Formella modeller av nätverkssystem
  - Insamling av data och mätvärden
  - Beräkning av optimala säkerhetsstrategier m.h.a förstärkningsinlärning
  - Implementera säkerhetsstrategier i *självlärande system*



Vår metod för att automatiskt beräkna säkerhetsstrategier

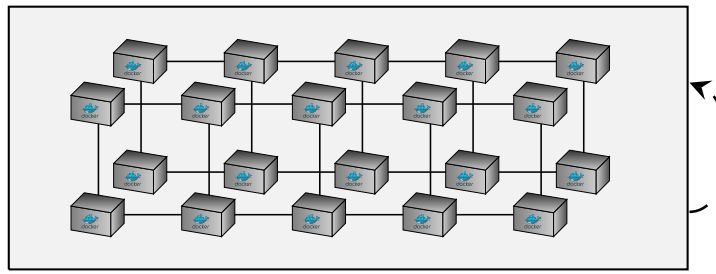


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

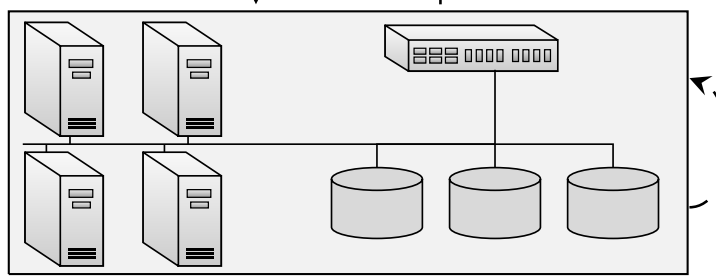


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

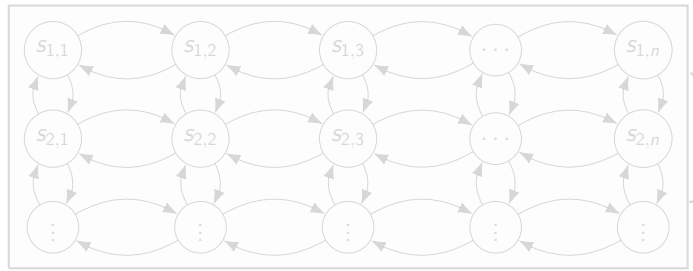
Selektiv replikering



IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

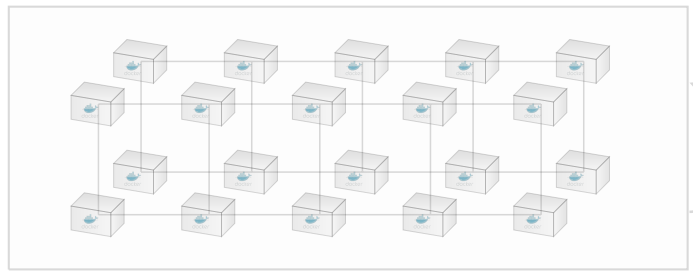


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



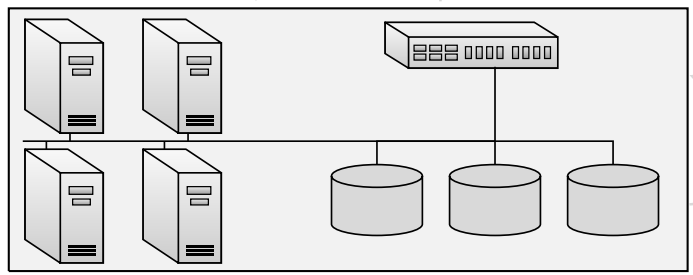
EMULERINGSSYSTEM

Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

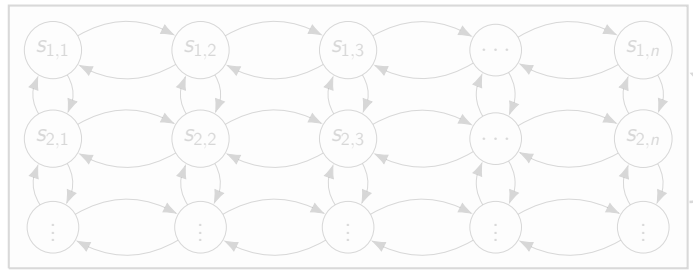
Selektiv replikering

IT INFRASTRUKTUR



Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

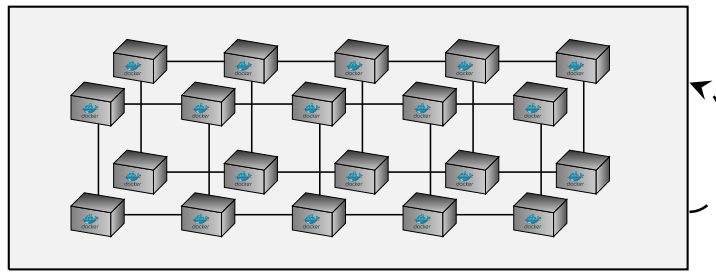


Simuleringssystem

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



Emuleringssystem

Utvärdering & skattning av modeller

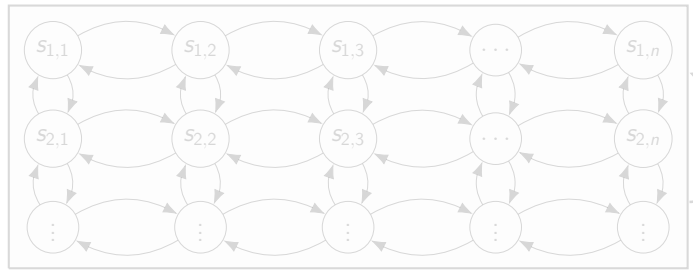
Implementation av säkerhetsstrategi π

Selektiv replikering

IT infrastruktur

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

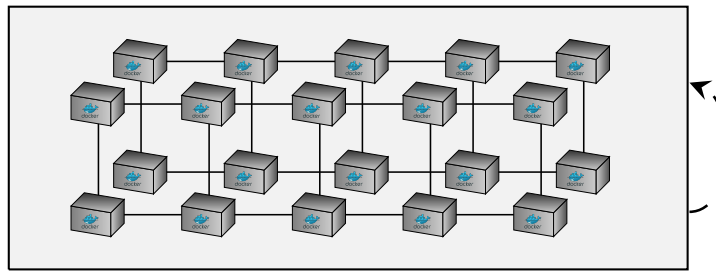


Simuleringssystem

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



Emuleringssystem

Utvärdering & skattning av modeller

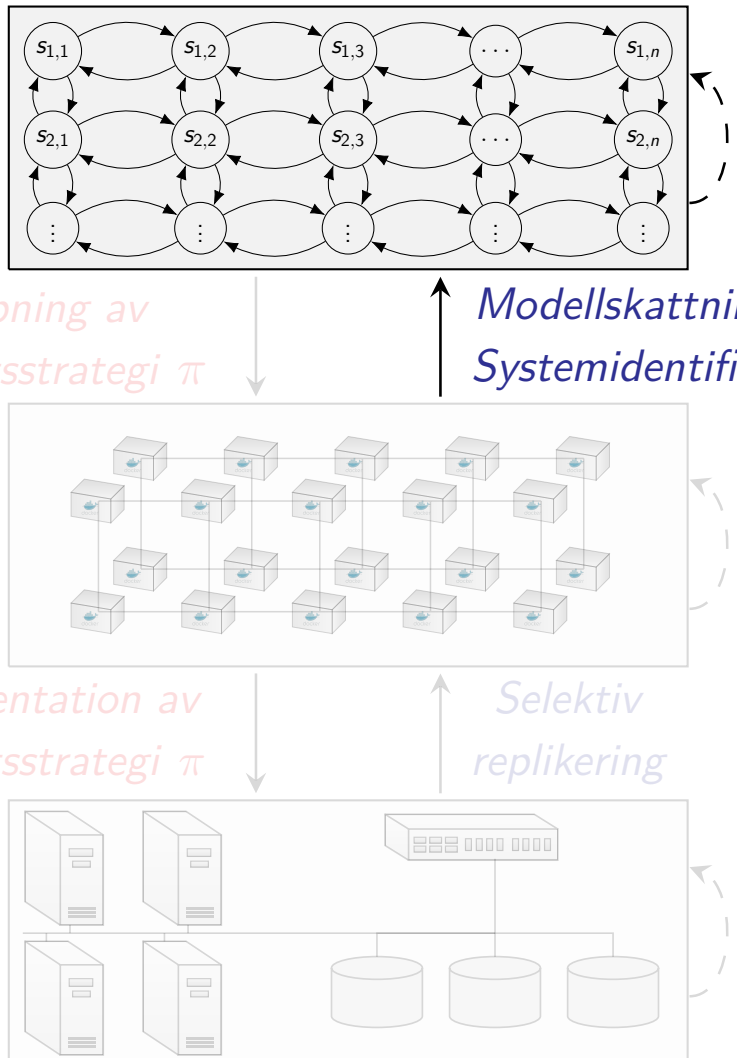
Implementation av säkerhetsstrategi π

Selektiv replikering

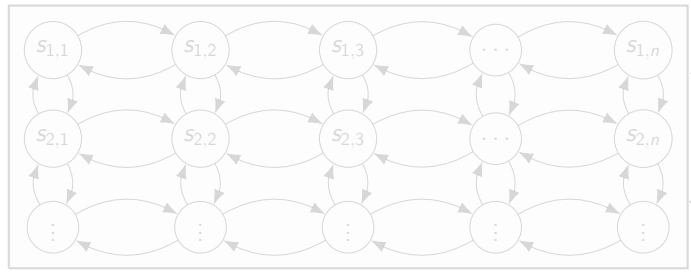
IT infrastruktur

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier



Vår metod för att automatiskt beräkna säkerhetsstrategier

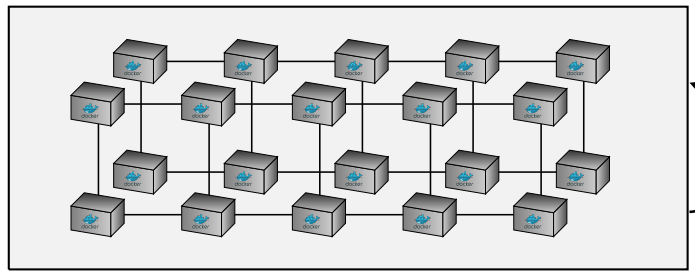


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering



EMULERINGSSYSTEM

Utvärdering & skattning av modeller

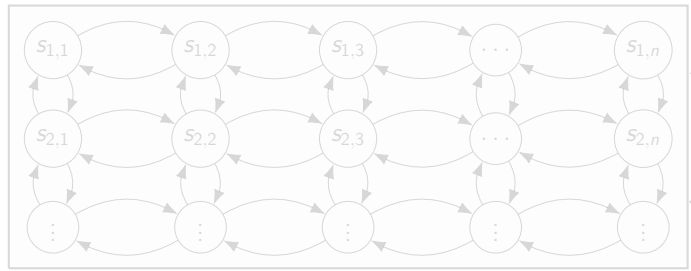
Implementation av säkerhetsstrategi π

Selektiv replikering

IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

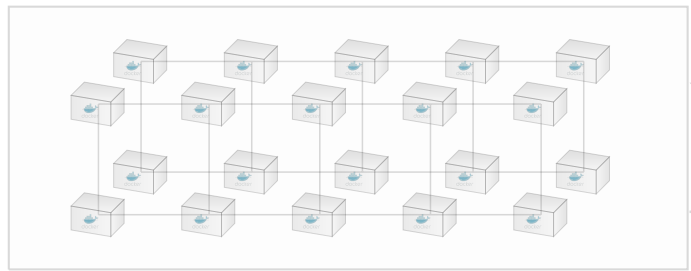


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

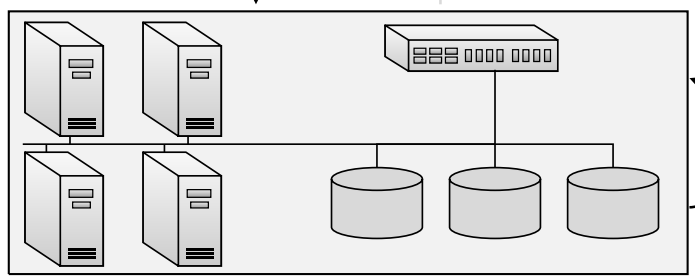


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

Implementation av säkerhetsstrategi π

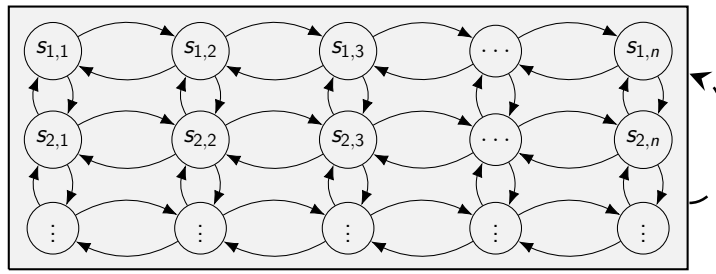
Selektiv replikering



IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system

Vår metod för att automatiskt beräkna säkerhetsstrategier

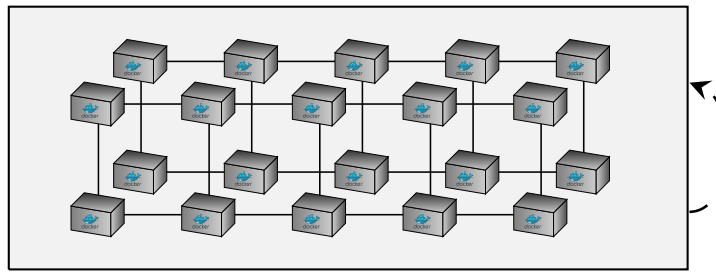


SIMULERINGSSYSTEM

Förstärkningsinlärning & optimering

Mappning av säkerhetsstrategi π

Modellskattning & Systemidentifiering

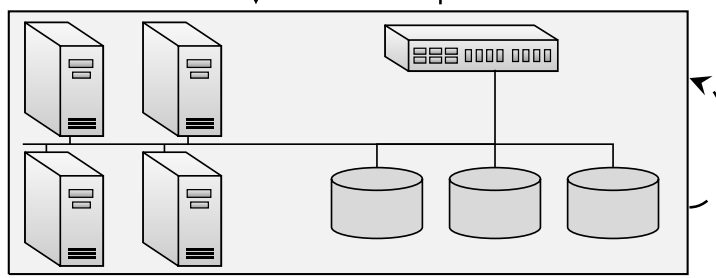


EMULERINGSSYSTEM

Utvärdering & skattning av modeller

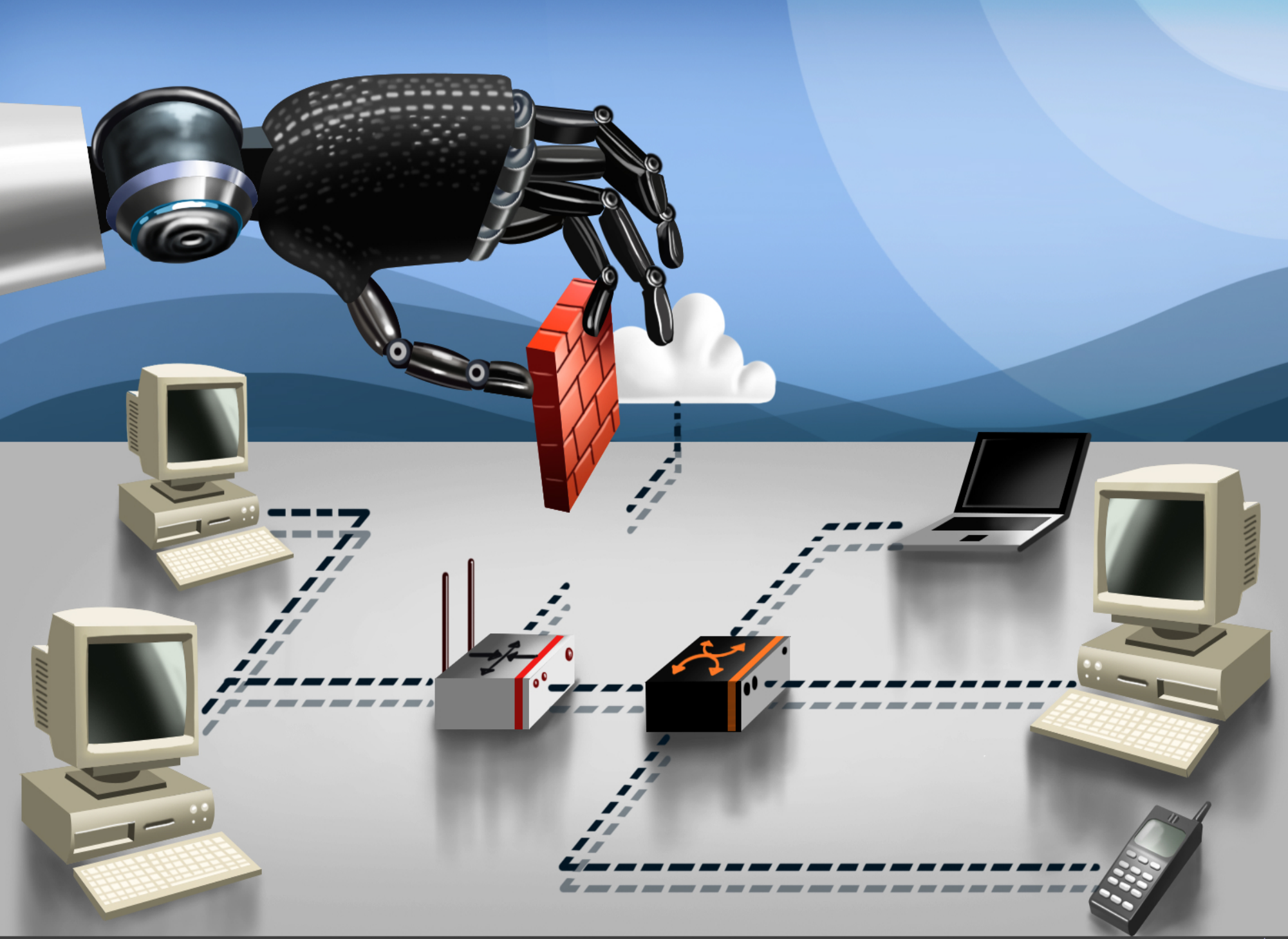
Implementation av säkerhetsstrategi π

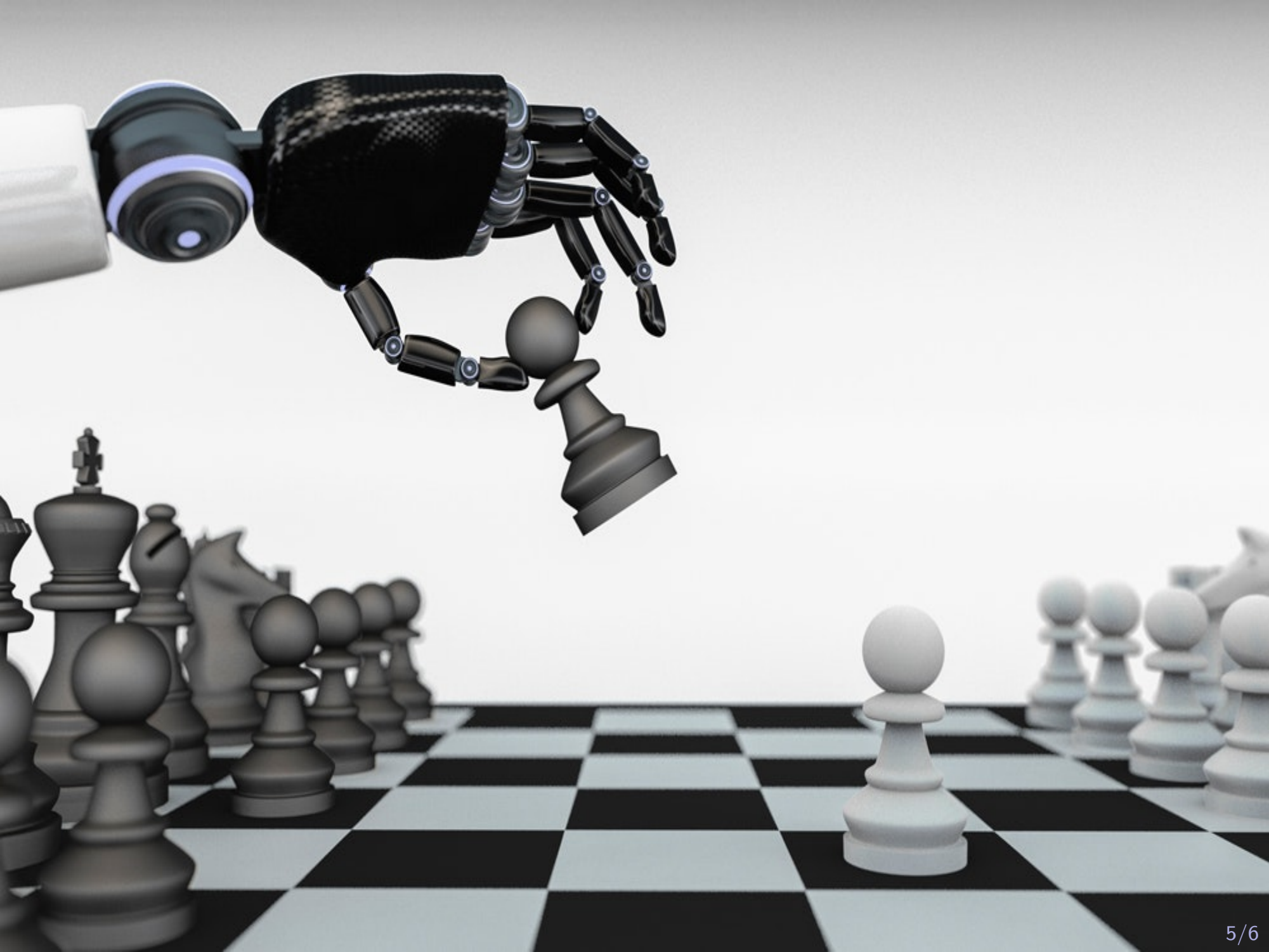
Selektiv replikering



IT INFRASTRUKTUR

Automatiserad säkerhet & självlärande system





Referenser

- *Finding Effective Security Strategies through Reinforcement Learning and Self-Play*[1]
  - **Preprint open access**: https://arxiv.org/abs/2009.08120
- *Learning Intrusion Prevention Policies through Optimal Stopping*[2]
  - **Preprint open access**: https://arxiv.org/pdf/2106.07160.pdf

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

[2] Kim Hammar and Rolf Stadler. *Learning Intrusion Prevention Policies through Optimal Stopping*. 2021. arXiv: 2106.07160 [cs.AI].