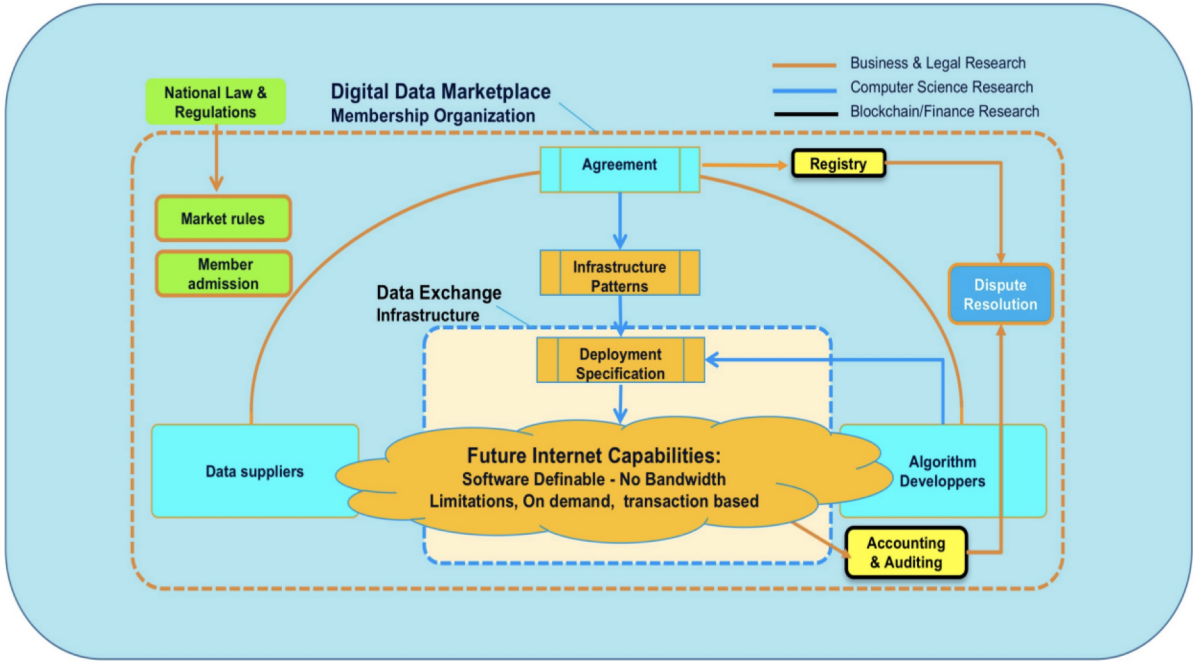
Risk level assessment for data exchange applications in Digital Data Marketplaces

Lu Zhang
MultiScale Networked Systems
University of Amsterdam

TKI DINALOG Dutch Institute for Advanced Logistics
NWO Netherlands Organisation for Scientific Research
Commit 2 Data connecting business and science
THALES BiZZdesign ciena evofenedex
AIRFRANCE KLM TRANSFIDES ORACLE
Gemeente Amsterdam
Universiteit van Amsterdam
TNO innovation for life

Digital Data Marketplaces (DDMs)

- **DDM** is a distributed **data trading platform** that supports data and/or compute asset sharing and federation among consortium members to achieve a common goal



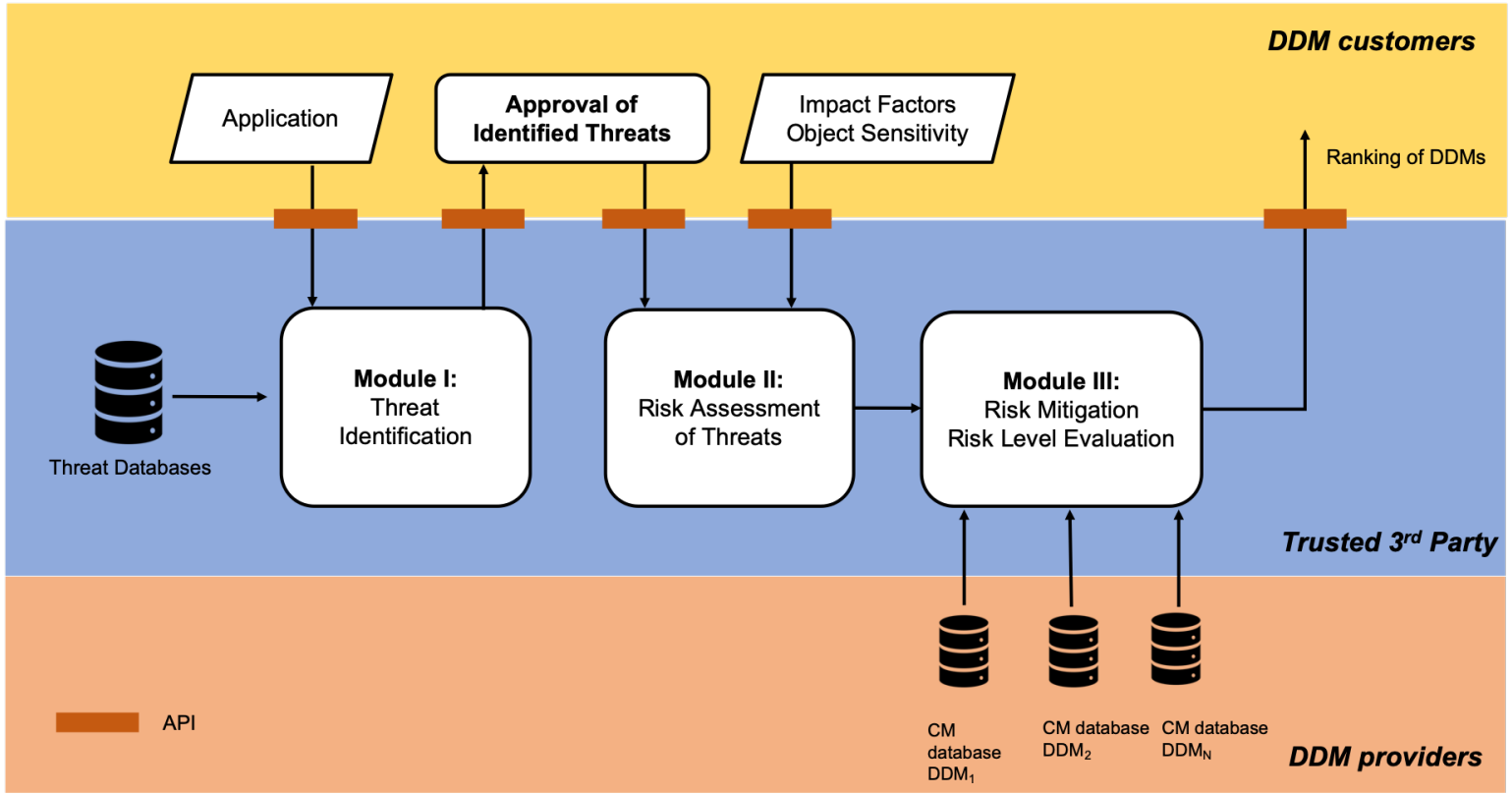
- Project **DL4LD** aims to facilitate trustworthy data sharing for a particular purpose with Digital Data Marketplace (DDM) concepts

- Security and sovereignty are top concerns in data federation applications

  How to allow DDM customers to choose an optimal DDM infrastructure with minimum **risk** for their applications?

A risk assessment system for DDMs

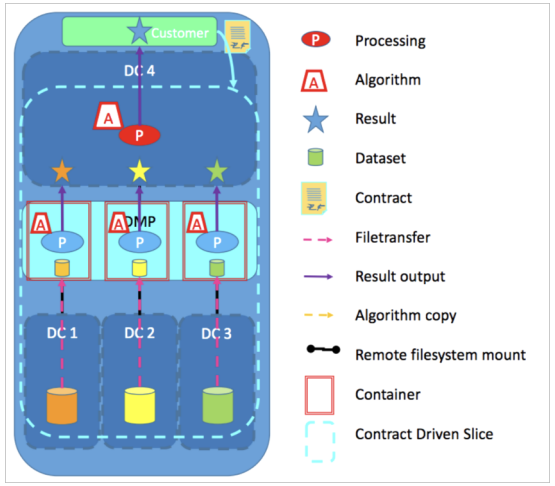
- Collaborative
- Application-based
- Robust
- Risk analysis-driven



Modified Microsoft STRIDE/DREAD model

Risk Attributes	Low (0)	Medium (5)	High(10)
Damage Potential (DP)	Depending on sensitivity value of Data Object, Compute Object and Result Object (Low, Medium, High)		
Accessibility (AC)	Only by consortium party member	By involving party e.g. 3rd party	By outsiders
Skill Level (SL)	Advanced skills	Malware existing in Internet or using attack tools	Simple tools
Affected Users (AU)	One party member	Partial party members	All party members
Intrusion Detectability (ID)	Detectable without monitoring	Detectable by monitoring	Very hard to detect by monitoring

Threat List (Approved)
Not-trustable computing env
Eavesdropping
Malicious code: high result correlation
Man-in-the-middle
Container runtime escape
Data loss: Physical attack
Dos on other containers



- Redefine risk attributes to address the concerns of DDM security assessment
  - Importance of monitoring
  - Potential trust among party members
- Risk attributes to estimate the probability of an exploitation of a vunerability

$$rs(t_i) = \frac{1}{5}(DP_{t_i} + AC_{t_i} + SL_{t_i} + AU_{t_i} + ID_{t_i})$$

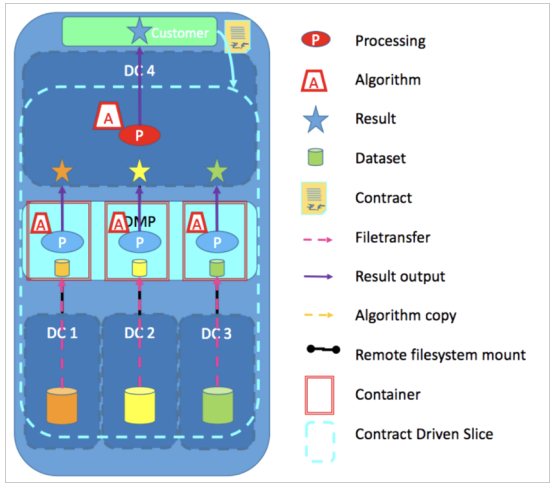
- Risk Ratio of each threat

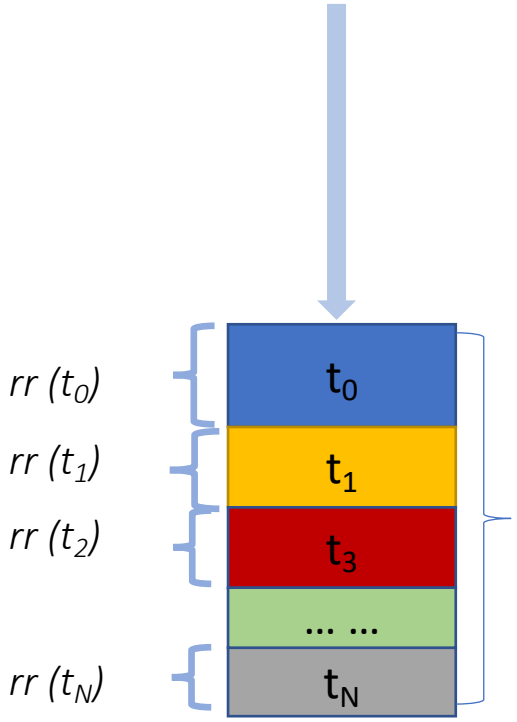
$$\mathrm{rr}(t_i) = \frac{rs(t_i)}{\sum_{t_i \in T} rs(t_i)}$$

Modified Microsoft STRIDE/DREAD model

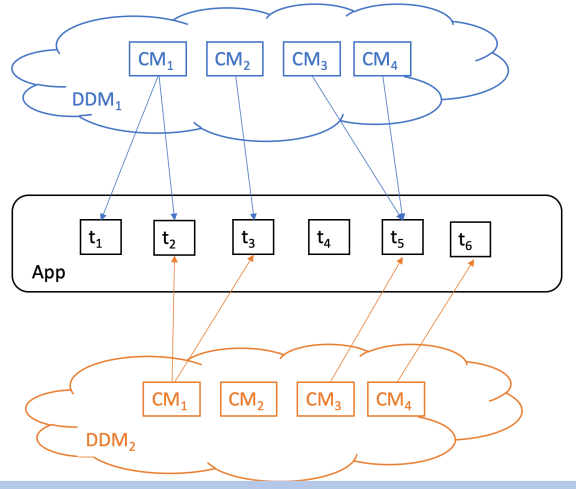
Risk Attributes	Low (0)	Medium (5)	High(10)
Damage Potential (DP)	Depending on sensitivity value of Data Object, Compute Object and Result Object (Low, Medium, High)		
Accessibility (AC)	Only by consortium party member	By involving party e.g. 3rd party	By outsiders
Skill Level (SL)	Advanced skills	Malware existing in Internet or using attack tools	Simple tools
Affected Users (AU)	One party member	Partial party members	All party members
Intrusion Detectability (ID)	Detectable without monitoring	Detectable by monitoring	Very hard to detect by monitoring

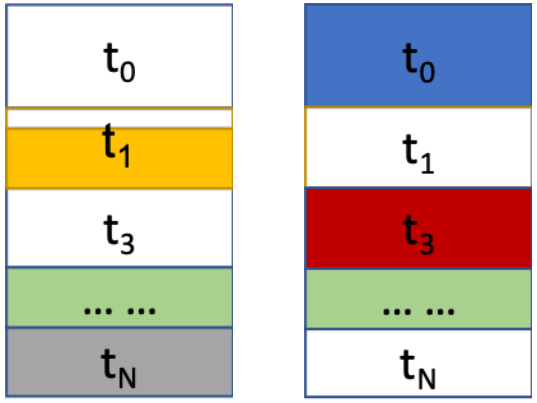
Threat List (Approved)
Not-trustable computing env
Eavesdropping
Malicious code: high result correlation
Man-in-the-middle
Container runtime escape
Data loss: Physical attack
Dos on other containers





100% risk





remaining risk (DDM_1) remaining risk (DDM_2)

Subjective choices of risk attributes values

- Use numeric values to represent qualitative levels
- Define a value vector $\overrightarrow{v_i}$ for numeric representations
  - $\overrightarrow{v_i} = [0, 5, 10]$ for original Microsoft model
- Define a metric ***Spreading Level*** to characterize the physical meaning of a value vector

$$SL(\overrightarrow{v_i}) = (v_{i,2} - v_{i,1}) - (v_{i,3} - v_{i,2}) \text{ with } \overrightarrow{v_i} = [v_{i,1}, v_{i,2}, v_{i,3}]$$

- The choices of value vectors with similar physical meaning turn to be subjective

Risk Attributes	Low (0)	Medium (5)	High(10)
Damage Potential (DP)	Depending on sensitivity value of Data Object, Compute Object and Result Object (Low, Medium, High)		
Accessibility (AC)	Only by consortium party member	By involving party e.g. 3rd party	By outsiders
Skill Level (SL)	Advanced skills	Malware existing in Internet or using attack tools	Simple tools
Affected Users (AU)	One party member	Partial party members	All party members
Intrusion Detectability (ID)	Detectable without monitoring	Detectable by monitoring	Very hard to detect by monitoring

System stability and resolution

- System Stability
  - Kendall's Tau τ
    - Similarity of two severity rankings of N threats with different value vectors
    - $\tau = \frac{\#\ concordant\ pairs - \#\ discordant\ pairs}{\binom{N}{2}}$
  - Normalized Mean Square Error (NMSE)
    - Variance of absolute values of *risk ratios* RR_x *and* RR_y *of N* threats with different value vectors
    - $NMSE(RR_x, RR_y) = \frac{1}{N}\sum_N \frac{(rr_i^{(x)} - rr_i^{(y)})^2}{\overline{RR_x}\cdot\overline{RR_y}}$, *with* $\overline{RR_x} = \frac{1}{N}\sum_N rr_i^{(x)}$
- System resolution
  - Granularity: the total number of unique values of risk scores for a given threat set

Experiment design

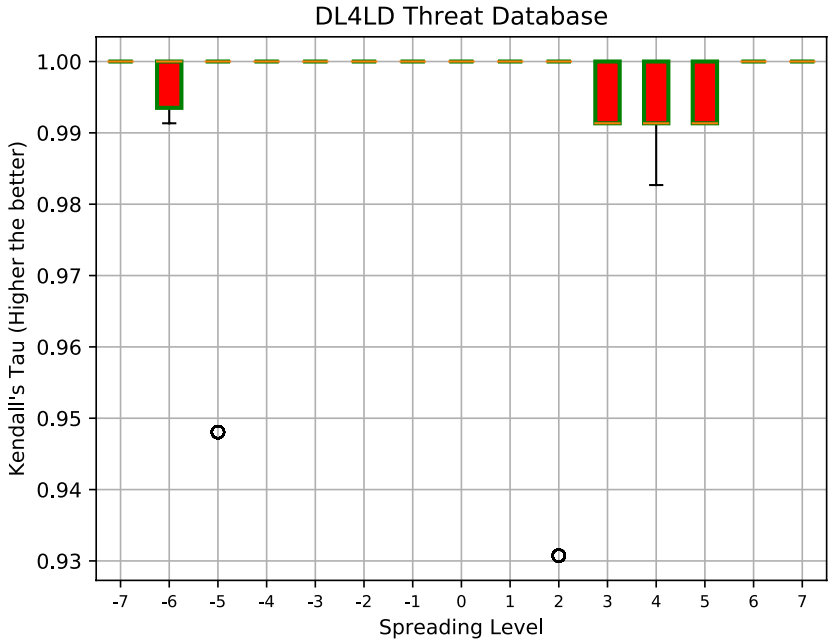
- Evaluate the fluctuations of ***risk ratios*** among ***value vectors*** of similar physical meaning
  - Group value vectors with SL
  - Within each equal SL cluster, compute mutual variance

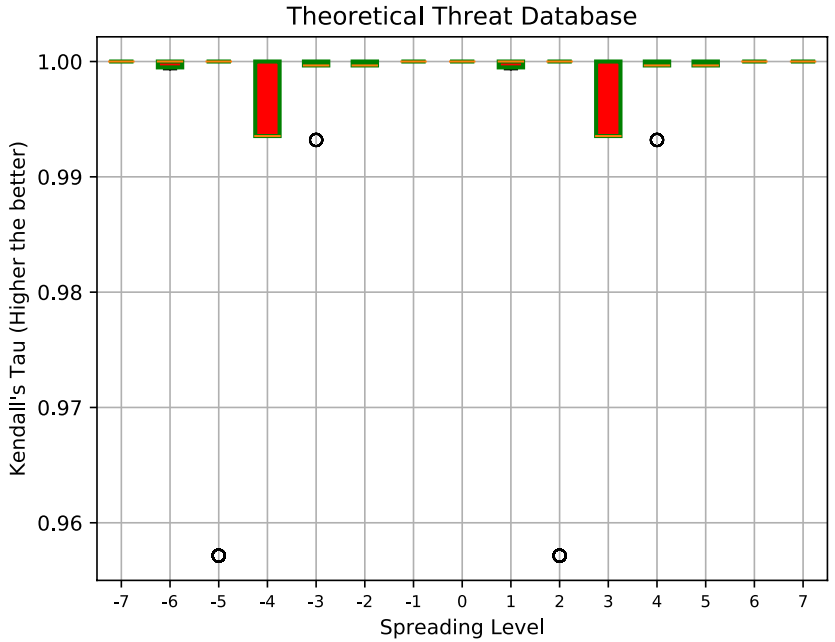
- Value Vector Set $V_{total} = \{\overrightarrow{v_1}, \overrightarrow{v_2}, \overrightarrow{v_3}, \dots\dots, \overrightarrow{v_i}\}$
  - $\overrightarrow{v_i} = [v_{i,1}, v_{i,2}, v_{i,3}]$ with $v_{i,j} \epsilon \{0, 1, \dots\dots, 10\}$

- Two Threat Databases
  - Theoretical Threat Database
  - DL4LD Threat Database
    - Threat Modeling of Archetypes

ThreatName	Stage	Category	Archetype	DP	AC	SL	AU	D
IP spoofing_	II	S	ALL	SO	H	M	H	M
Identity spoofing: via remote data access_	III	S	IV, V, VII	SO	H	L	M	H
Insecure data deletion_	III	ID	ALL	SO	M	L	M	H
Malicious compute: Data Disclosure_	III	ID	ALL	SO	L	H	H	M
Unauthorized disclosure: Eavesdropping_	II	ID	ALL	SO	H	H	M	H
Weak Access Control	I	ID	ALL	SO	H	H	L	H
Malicious compute: high result correlation_	III	ID	III	SO	L	H	H	M
Encryption Keys Leakage during exchange:	II	ID	ALL	TOP	H	L	H	H
Cross-tenant Side Channel Attack_	III	ID	IV, V, VI, VII	SO	M	L	H	H
Management Interface Compromise_	I, III	ID, T	IV, V, VII	SO	M	M	M	M
Isolation Failure: Poorly separated container traffic_	III	ID	VII	SO	L	L	H	H
Isolation Failure: Cross vm/container attack_	III	ID	IV, V, VI, VII	SO	M	L	H	H

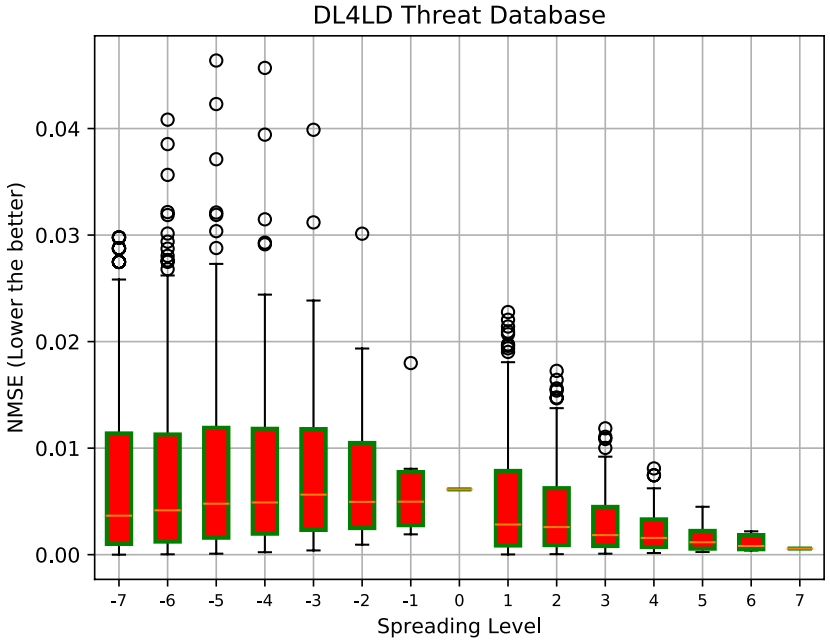
Result analysis of system stability - Kendall's Tau

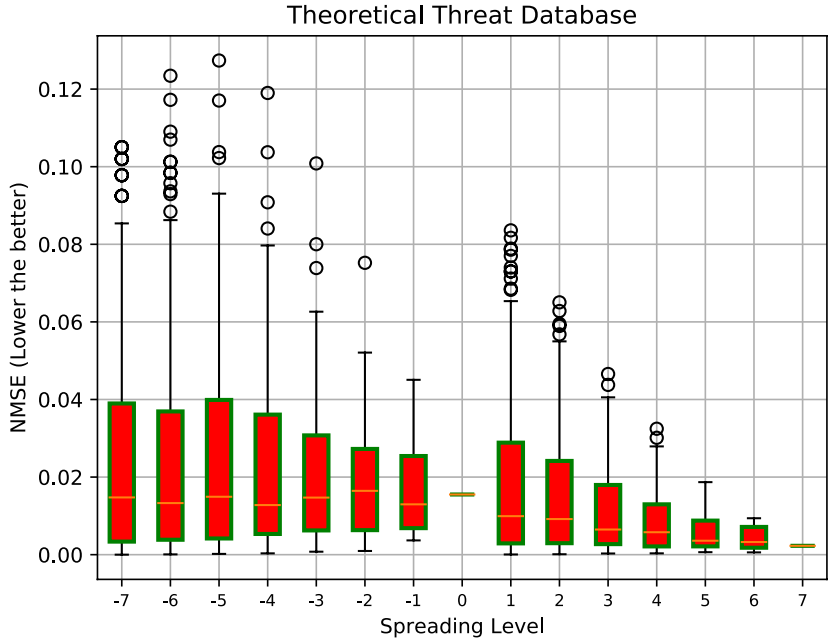




- The severity ranking is totally robust for evenly-distributed value vectors for all real-world threat databases
- In general, the influence is subtle due to subject choices of value vectors for threat severity rankings

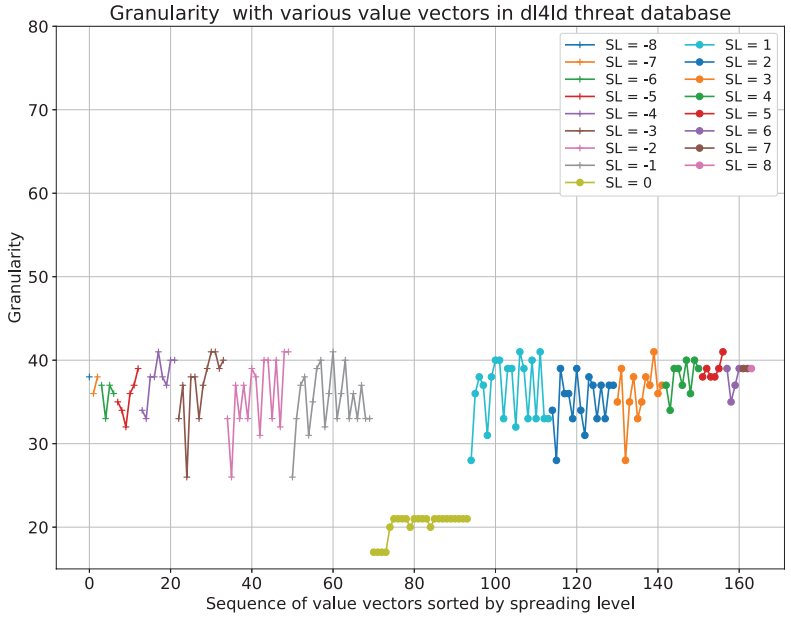
Result analysis of system stability - NMSE

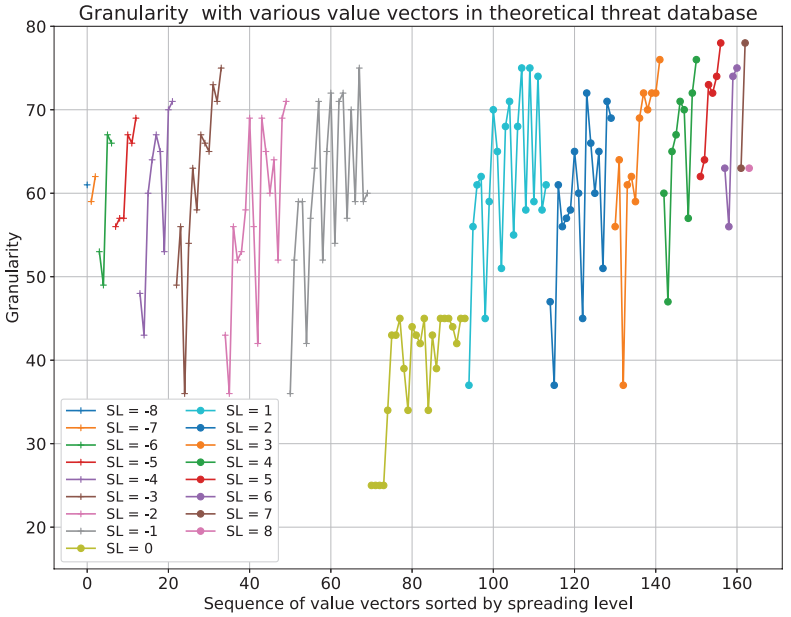




- Our methodology performs much better for the DL4LD use case compared to the theoretical threat database
- Very subtle chance to vary the final DDM rankings due to subject choices of parameters

Result analysis of system resolution - Granularity





- The values of Granularity fluctuates for value vectors of same SL
  - It is recommended for users to choose a value vector with relatively high Granularity and to avoid those with very low resolution.
- The DL4LD use case performs very well regarding to provided resolution

Conclusions and future work

- ✓ We proposed a system to quantitatively assess the risk level of applications in DDMs
  - Capture the dynamic features
  - Focus on specific concerns of DDM applications
- ✓ We validated the stability and resolution of our system, specifically for the DL4LD use case
  - Subjective choices of users have very subtle influence on the provided DDM rankings of the system
- ➢ Further improve the risk assessment system to be adaptive
  - Real time risk level VS applied countermeasures

THANK YOU AND ANY QUESTIONS?

www.dl4ld.nl
www.dl4ld.net