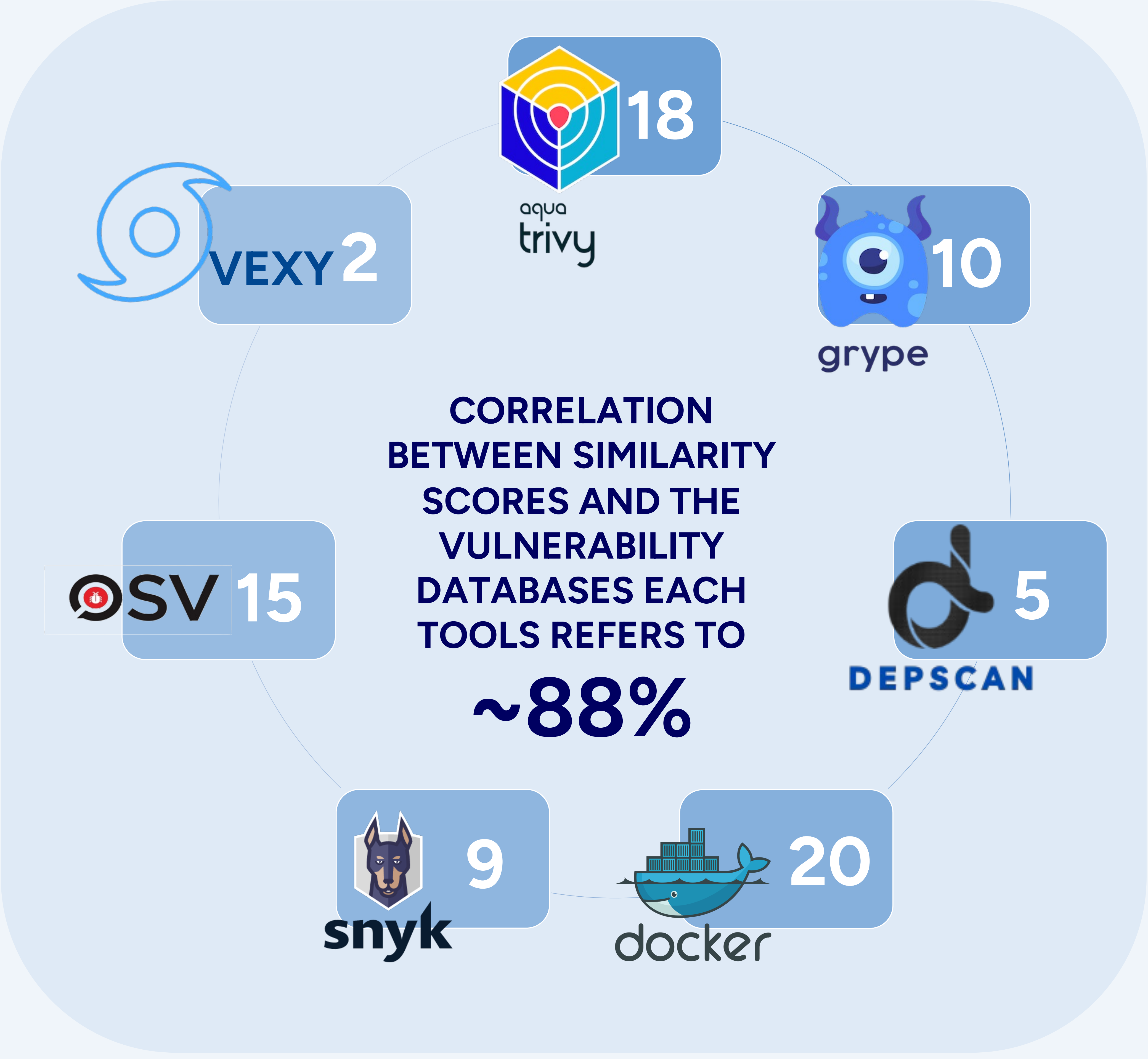
Vexed by VEX tools: Consistency evaluation of container vulnerability scanners

Yekatierina Churakova (yekchu@kth.se)
Supervised by: **Mathias Ekstedt** (mekstedt@kth.se)



CORRELATION BETWEEN SIMILARITY SCORES AND THE VULNERABILITY DATABASES EACH TOOLS REFERS TO

~88%

Number of vulnerability databases for each tool

Key findings

The tools produce very inconsistent reports. In general they report more different findings than same findings – for the same container

The construction of reliable ground truth dataset is not meaningful for the study due to low similarity in the reports.

Naming problem affects the similarity.

In our experiments, we observed an increase in similarity scores for reports based on a single vulnerability identifier.

Similarity scores based on Jaccard index

Subset	Trivy	Grype	DepScan	Scout	Snyk	OSV	Vexy
Trivy	1	**0.694**	0.160	0.329	0.379	0.059	0.018
Grype	**0.694**	1	0.155	0.304	0.355	0.004	0
DepScan	0.160	0.155	1	0.062	0.118	0.010	0.003
Scout	0.379	0.304	0.062	1	0.332	0.129	0.041
Snyk	0.379	0.355	0.118	0.332	1	0.003	0
OSV	0.059	0.004	0.010	0.129	0.003	1	0.095
Vexy	0.018	0	0.003	0.041	0	0.095	1

Similarity scores for CVE-only vulnerabilities

Subset	Trivy	Grype	DepScan	Scout	Snyk	OSV	Vexy
Trivy	1	**0.76**	0.163	0.334	0.387	0.062	0.015
Grype	**0.76**	1	0.162	0.33	0.355	0.004	0
DepScan	0.163	0.162	1	0.062	0.118	0.007	0.04
Scout	0.334	0.33	0.062	1	0.34	0.126	0.03
Snyk	0.387	0.355	0.118	0.34	1	0.003	0
OSV	0.062	0.004	0.007	0.126	0.003	1	0.18
Vexy	0.015	0	0.004	0.03	0	0.18	1

Recommendation:

Use several tools with relatively low consistency (but not too low) to have a better coverage with minimum level of false-positives.

KTH ROYAL INSTITUTE OF TECHNOLOGY

April 2025