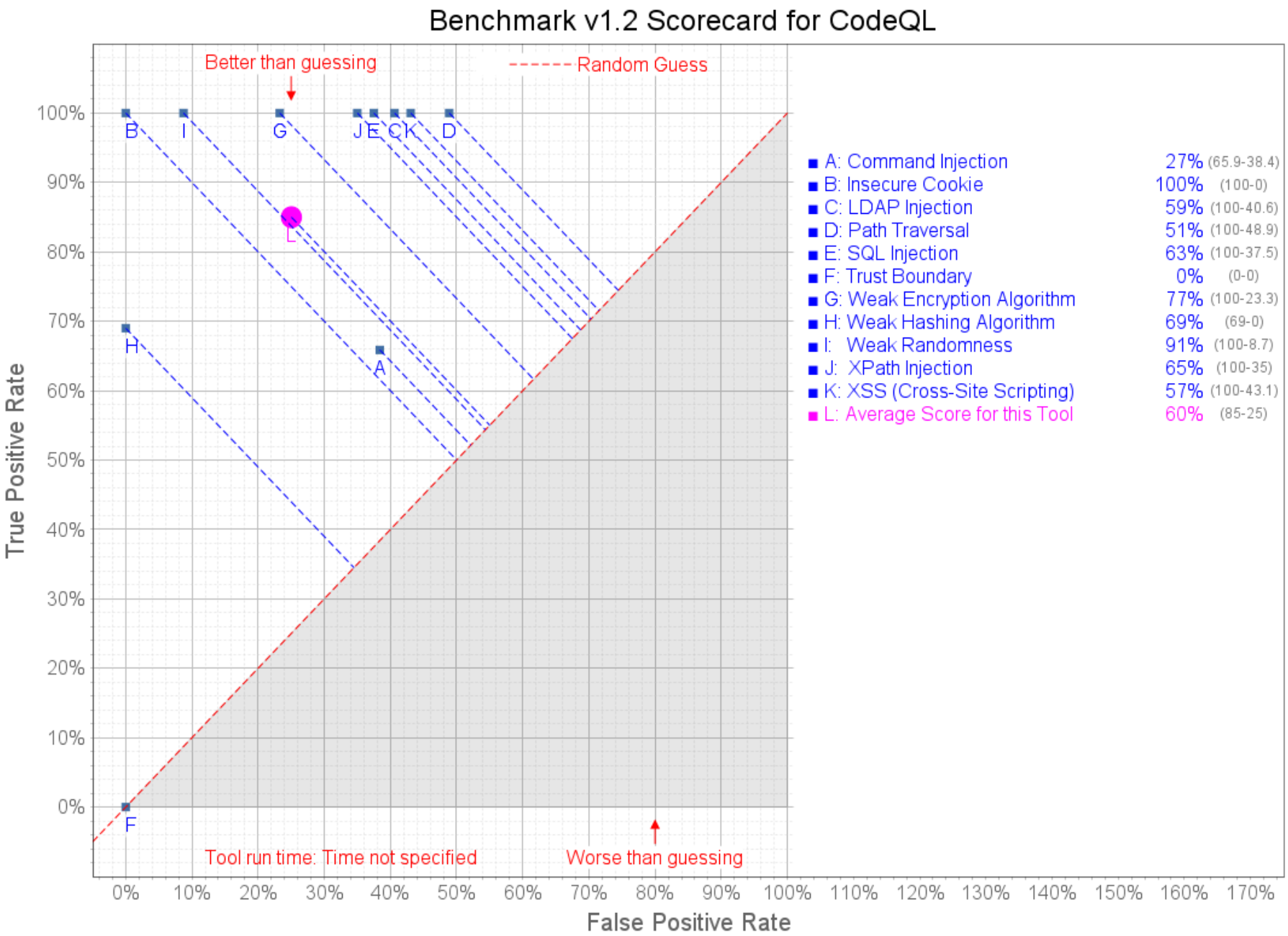
empty
empty

OWASP Benchmark Scorecard for CodeQL v2.17.1 (SAST)

OWASP Benchmark is a test suite designed to evaluate the speed, coverage, and accuracy of automated vulnerability detection tools. Without the ability to measure these tools, it is difficult to understand their strengths and weaknesses, and compare them to each other. OWASP Benchmark contains thousands of test cases that are fully runnable and exploitable. The following is the scorecard for the tool CodeQL against version 1.2 of OWASP Benchmark. It shows how well this tool finds true positives and avoids false positives in the OWASP Benchmark test cases.

For more information, please visit the OWASP Benchmark Project Site (https://owasp.org/www-project-benchmark/).



Legend: Each line of the scorecard in the upper right has a category, followed by this tool's overall score in that category. After each % is this tool's TPR - FPR used to calculate this score. See the table below for more precise scoring details.

Statistics

Tool elapsed analysis time	Time not specified
Tool overall score (0-100)	59.94%
Total test cases	2740
Download raw results	Actual Results (Benchmark_v1.2_Scorecard_for_CodeQL_v2.17.1.csv)

Detailed Results

Category	CWE #	TP	FN	TN	FP	Total	TPR	FPR	Score
Command Injection	78	83	43	77	48	251	65.87%	38.40%	27.47%
Insecure Cookie	614	36	0	31	0	67	100.00%	0.00%	100.00%
LDAP Injection	90	27	0	19	13	59	100.00%	40.62%	59.38%
Path Traversal	22	133	0	69	66	268	100.00%	48.89%	51.11%
SQL Injection	89	272	0	145	87	504	100.00%	37.50%	62.50%
Trust Boundary	501	0	83	43	0	126	0.00%	0.00%	0.00%
Weak Encryption Algorithm	327	130	0	89	27	246	100.00%	23.28%	76.72%
Weak Hashing Algorithm	328	89	40	107	0	236	68.99%	0.00%	68.99%
Weak Randomness	330	218	0	251	24	493	100.00%	8.73%	91.27%
XPath Injection	643	15	0	13	7	35	100.00%	35.00%	65.00%
XSS (Cross-Site Scripting)	79	246	0	119	90	455	100.00%	43.06%	56.94%

Totals	**1249**	**166**	**963**	**362**	**2740**			
Overall Results*						**84.99%**	**25.04%**	**59.94%**

*-The Overall Results are averages across all the vulnerability categories. You can't compute these averages by simply calculating the TPR and FPR rates using the values in the Totals row. If you did that, categories with larger number of tests would carry more weight than categories with less tests. The proper calculation of the Overall Results is to add up the values of each of these per row, then divide by the number of rows, which is how they are calculated.

Key

Common Weakness Enumeration (CWE)	The primary MITRE CWE number for this vulnerability category.
True Positive (TP)	Tests with real vulnerabilities that were correctly reported as vulnerable by the tool.
False Negative (FN)	Tests with real vulnerabilities that were not correctly reported as vulnerable by the tool.
True Negative (TN)	Tests with fake vulnerabilities that were correctly not reported as vulnerable by the tool.
False Positive (FP)	Tests with fake vulnerabilities that were incorrectly reported as vulnerable by the tool.
True Positive Rate (TPR) = TP / (TP + FN)	The rate at which the tool correctly reports real vulnerabilities. Also referred to as Recall, as defined at Wikipedia (https://en.wikipedia.org/wiki/Precision_and_recall).
False Positive Rate (FPR) = FP / (FP + TN)	The rate at which the tool incorrectly reports fake vulnerabilities as real.
Score = TPR - FPR	How good the tool is at finding true positives and avoiding false positives.