A Detailed results of the 12-hour experiments

The results of the 12-hour experiment are given in the following table (Table 1), in which we highlight the best results with dark gray and the second-best results with light gray to inspect the primary influence factors.

Table 1. Coverage results of the 12-hour experiment. The metrics B and S represent branch and statement coverages, respectively.

Programs		DFS	BFS	Lase	NS	GS#D	GS#B
Csharp2	B	700.0	700.0	700.0	699.7	673.0	693.0
	S	1373.0	1373.0	1373.0	1373.0	1316.0	1363.0
Jsijcc	B	1648.0	1611.0	1736.0	1728.0	1645.7	1684.0
	S	2799.7	2761.0	2889.0	2882.0	2749.7	2817.7
Gap	B	1208.0	1202.0	1287.0	1275.7	1188.7	1137.0
	S	1753.0	1738.7	1840.0	1828.0	1720.7	1664.0
Java	B	1845.0	1755.0	2151.7	2140.3	1837.3	1870.3
	S	3142.0	2811.7	3552.0	3472.7	3166.7	3074.3
C6	B	590.0	597.0	668.0	598.0	621.7	621.7
	S	1234.0	1261.0	1323.0	1264.0	1278.0	1298.7
Feel	B	833.0	973.0	1015.3	1010.3	835.3	980.7
	S	1280.0	1495.0	1539.3	1534.3	1282.0	1499.7
R	B	1116.0	1120.7	1166.3	1122.0	1158.7	1164.3
	S	2169.0	2190.3	2295.7	2188.0	2260.3	2287.7
Php	B	898.0	891.0	907.3	907.0	893.3	897.3
	S	2101.0	2056.0	2132.0	2127.3	2079.7	2099.0
Java6	B	714.0	770.3	770.0	765.3	711.0	772.7
	S	1491.7	1631.7	1640.0	1626.0	1479.0	1641.0
OaJava	B	1783.0	1793.7	2220.3	2221.0	1900.3	1892.7
	S	3527.0	3378.7	4209.7	4216.0	3753.7	3596.3
Sixpath	B	1221.0	1190.0	1099.0	1137.0	1117.7	1117.7
	S	1854.0	1844.0	1768.3	1810.7	1787.3	1793.0
Selfdef7	B	1030.3	1071.3	1087.7	1072.7	1047.3	1057.7
	S	1720.7	1812.7	1828.0	1812.0	1764.0	1803.3
Java1	B	1411.0	1518.7	1815.3	1809.3	1522.3	1624.0
	S	2018.7	2064.7	2423.7	2416.0	2142.7	2213.0
Antlrjava	B	755.0	878.0	835.0	848.7	803.3	875.7
	S	1976.0	2334.0	2238.3	2287.3	2124.7	2332.0
C15	B	607.0	614.0	608.0	610.0	590.0	615.3
	S	1327.0	1343.0	1330.0	1336.0	1263.3	1343.7

Note that for Csharp2, all methods are able to fully explore the execution paths for the given inputs, resulting in almost identical coverage. For Feel, we observe that the coverage achieved after 12 hours is actually worse than that after just 1 hour. This is because, in the 12-hour experiment, we allocated 1 hour to symbolically execute the tokenization code in the first phase. While this first phase discovers more tokens, it may lead to reduced grammar rule coverage in the second phase, as a larger set of candidate tokens at each position may dilute the exploration of specific grammar rules.

Figure 1 shows the comparison of coverage increases for **Lase**, **NS**, and **GS#D** relative to **DFS** on the functionality code of different compilers, while Figure 2 display the results of **Lase**, **NS**, and **GS#B** relative to **BFS**. _B means branch coverage and _S means statement coverage. For Java6, Php and Jsijcc, **NS** performs worse than the baselines because the new search strategy tends to cover grammar rules more quickly, without ensuring valid input generation.

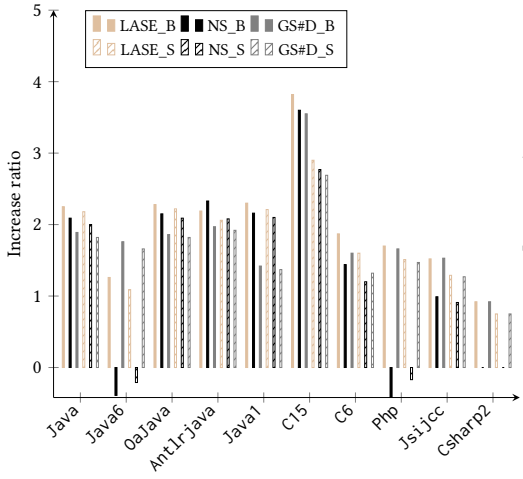

Fig. 1. Relative increase to DFS (12 hours)

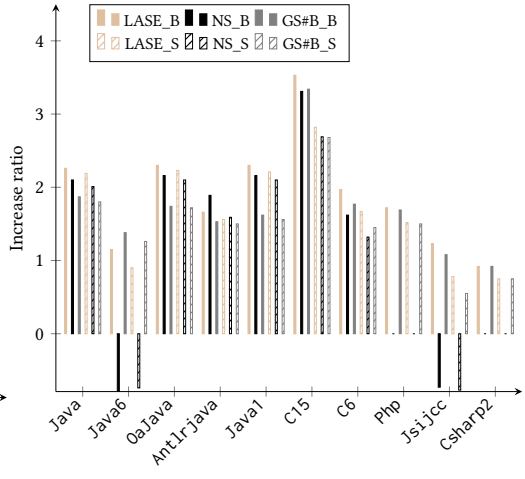

Fig. 2. Relative increase to BFS (12 hours)

B Detailed results of the scalability experiment

The detailed coverage statistics for the synthetic grammar (whose parser program has 115,692 lines of code) under the 1-hour and 12-hour settings are reported in Table 2 and Table 3, respectively. On this benchmark, Lase consistently outperforms all other approaches. The proposed search strategy and grammar synthesis further enhance the effectiveness of symbolic execution. Notably, grammar synthesis plays a more critical role, as the grammar rules in the synthetic grammar share a higher degree of similarity.

Table 2. Coverage results of the synthetic grammar (1 hour). The metrics B and S represent branch and statement coverage, respectively.

Program		DFS	BFS	Lase	NS	GS#D	GS#B
SyntheticG	B	970.3	1262.3	2104.3	1413.0	1788.7	2016.0
	S	3218.0	4262.7	8053.3	4943.3	6931.3	7744.7

Table 3. Coverage results of the synthetic grammar (12 hours). The metrics B and S represent branch and statement coverage, respectively.

Program		DFS	BFS	Lase	NS	GS#D	GS#B
SyntheticG	B	1269.0	1485.3	2728.0	1581.0	2112.7	2317.0
	S	4355.0	4981.3	10078.3	5467.0	7883.3	8570.0