Online Appendix for "FCEVAL: An Effective and Quantitative Platform for Evaluating Fuzzer Combinations Fairly and Easily"

ABSTRACT

1. Introduction

This document provides additional material for the paper "FCEVAL: An Effective and Quantitative Platform for Evaluating Fuzzer Combinations Fairly and Easily".

2. Evaluation for FCC and FCD

We offer detailed evaluation results here for extensive observations.All experiment settings are identical to the evaluation for FCA and FCB except for benchmarks. We choose 6 programs from Binutils for assessment.The evaluation results for FCC and FCD would be shown in the following tables and figures.

ORCID(s):

Table 1
Total number of unique branches covered by FCC with the two sharing policies of test cases.

	nm	objdump	readelf	size	strings	strip
$ENFUZZ^1$	5372	2656	7242	1664	1336	2711
$FCEVAL^2$	5403	2865	7291	1676	1357	2727
p-value	0.74	0.22	<0.01	0.54	0.96	0.49
$\hat{A}12^3$	0.55	0.67	0.87	0.59	0.49	0.60

$ENFUZZ^1$: the total number of unique branches covered with POLICY_ENFUZZ.
$FCEVAL^2$: the total number of unique branches covered with POLICY_FCEVAL.
$\hat{A}12^3$:with POLICY_ENFUZZ as the baseline.

Table 2
Total number of unique bugs found by FCC with the two sharing policies of test cases.

	nm	objdump	readelf	size	strings	strip
$ENFUZZ^1$	569	3	0	0	2	14
$FCEVAL^2$	655	2	0	2	2	29
p-value	0.85	1	1	0.47	1	0.63
$\hat{A}12^3$	0.53	0.51	0.50	0.60	0.55	0.44

$ENFUZZ^1$: the total number of unique bugs discovered by FCC with POLICY_ENFUZZ.
$FCEVAL^2$: the total number of unique bugs found by FCC with POLICY_FCEVAL.
$\hat{A}12^3$: with POLICY_ENFUZZ as the baseline.

Table 3
The detail of global branch coverage by FCC and FCD with POLICY_FCEVAL.

Benchmark	FC	Total	Avg	p-value	$\hat{A}12$	Smax	Time (Minute)	Complementarity
nm	FCD	4305	771	-	-	1285	1432	1940.22
nm	FCC	6326	1527	<0.01	0.99	1285	161	3658.23
objdump	FCD	2953	652	-	-	767	1438	1581.12
objdump	FCC	3445	663	0.42	0.98	767	158	1706.07
readelf	FCD	5902	1602	-	-	1887	1437	3709.92
readelf	FCC	7778	2834	<0.01	1.0	1887	54	5981.03
size	FCD	1820	444	-	-	502	1439	978.43
size	FCC	1990	530	<0.01	1.0	502	519	1162.40
strings	FCD	1405	368	-	-	396	1423	781.98
strings	FCC	1653	450	<0.01	1.0	396	141	954.92
strip	FCD	2717	627	-	-	784	1400	1449.13
strip	FCC	3331	830	<0.01	0.58	784	511	1871.07

FC: fuzzer combination. Total: total unique branches covered. Avg: average number of covered branches.
$\hat{A}12$: with FCD as the baseline. Smax: smaller one of fuzzer combinations' maximum branches.

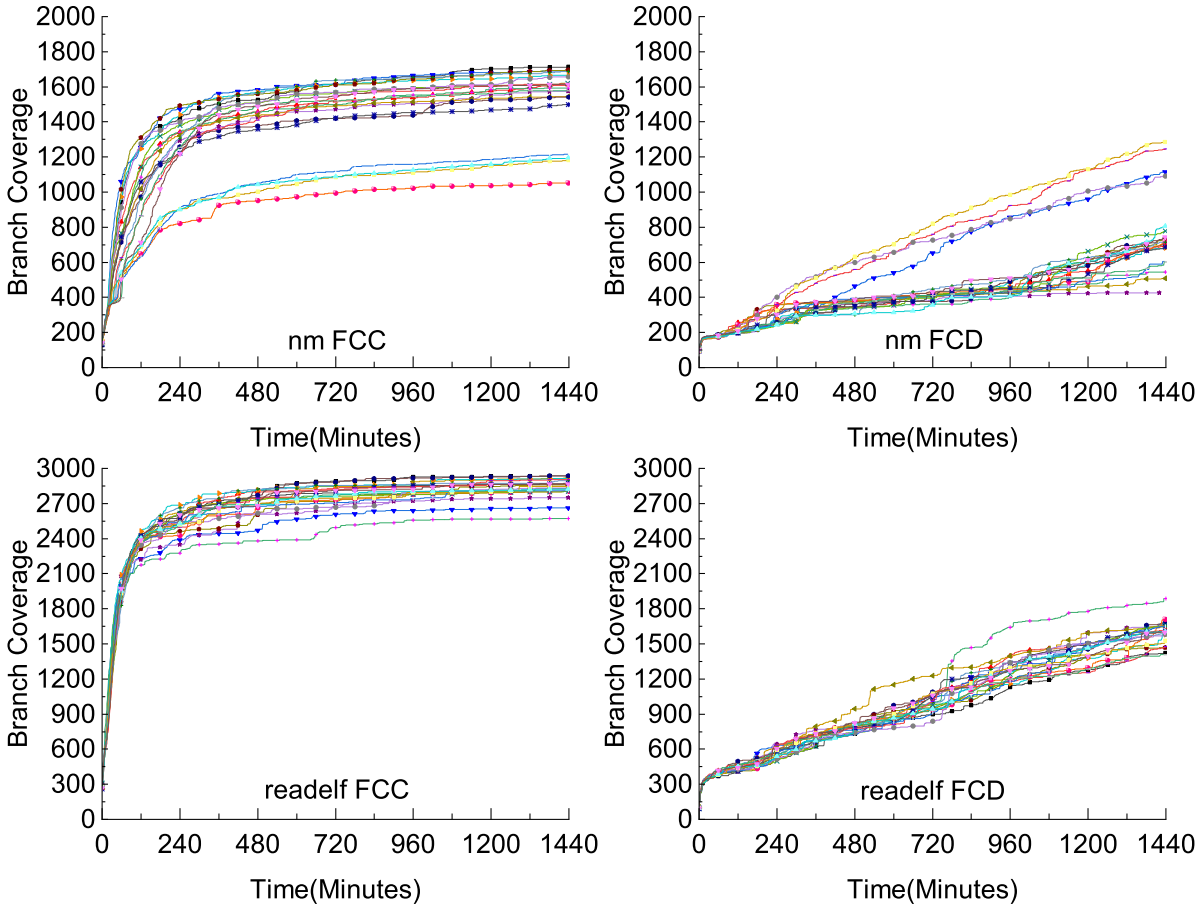

Figure 1: The total number of branches covered by FCC and FCD on nm and readelf over time.

Table 4
The detail of unique bugs found by FCC and FCD with POLICY_FCEVAL.

Benchmark	FC	Total	Avg	p-value	$\hat{A}12$	Faster
nm	FCC	175	21.75	-	-	22
nm	FCD	979	186.4	<0.01	0.98	972
objdump	FCC	2	1.8	-	-	0
objdump	FCD	5	1.35	0.13	0.58	5
readelf	FCC	0	0.0	-	-	0
readelf	FCD	0	0.0	1	1.0	0
size	FCC	2	0.1	-	-	0
size	FCD	2	0.1	1	0.99	2
strings	FCC	0	0.0	-	-	0
strings	FCD	2	0.2	0.49	1.0	2
strip	FCC	13	3.35	-	-	9
strip	FCD	32	7.2	<0.01	1.0	28

FC: fuzzer combination. Total: total number of unique bugs.
$\hat{A}12$: with FCB as the baseline.
Avg: average number of bugs found in repeated experiments.
Faster: number of bugs found faster by FCC or FCD.

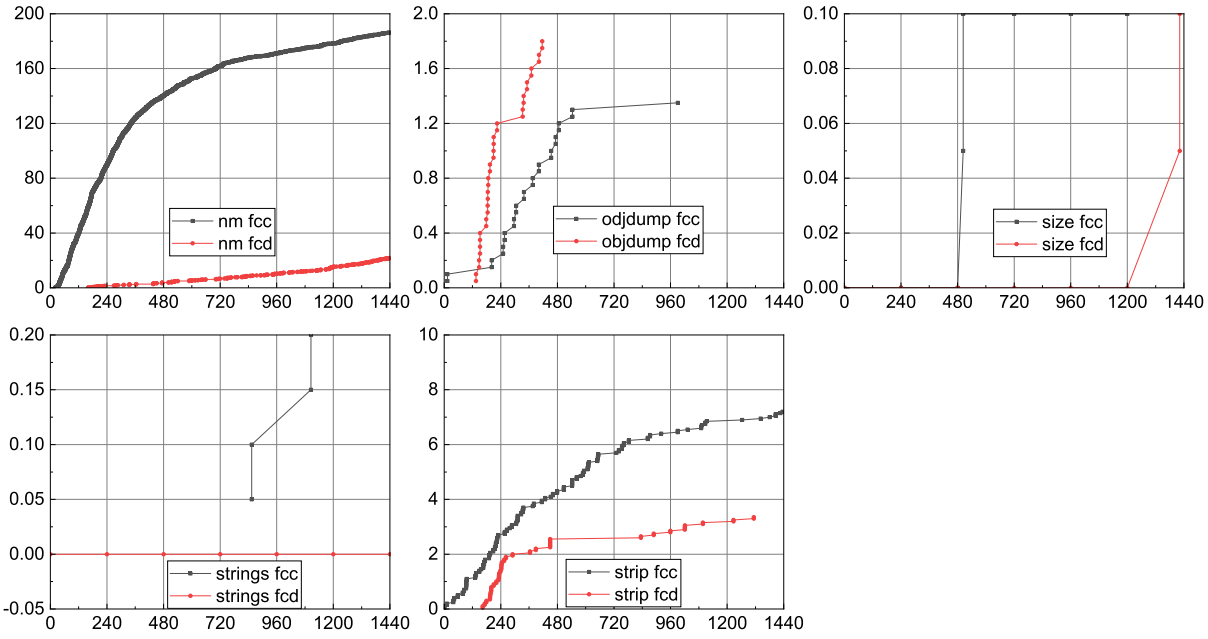

Figure 2: The average number of unique bugs found by FCC and FCD over time with the unit of minute. "nm fcc" denotes that of FCC on nm.

Table 5
Spearman r_s between coverage and bugs grouped by both fuzzer combination and benchmark.

FC	nm	objdump	readelf	size	strings	strip
FCC	0.7	0.78	-	0.3	-0.09	-0.06
FCD	0.79	0.52	-	0.14	-	0.26

FC: fuzzer combination.