POWER : Program Option-Aware Fuzzer for High Bug Detection Ability (ICST 2022)

Ahcheong Lee (KAIST, South Korea)
Irfan Ariq (KAIST)
Yunho Kim (Hanyang Univ. , South Korea)
Moonzoo Kim (KAIST)

ICST 2023 Ireland

POWERUP: Program Option-Aware Interleaving Fuzzing Platform for High Bug Detection
TBD

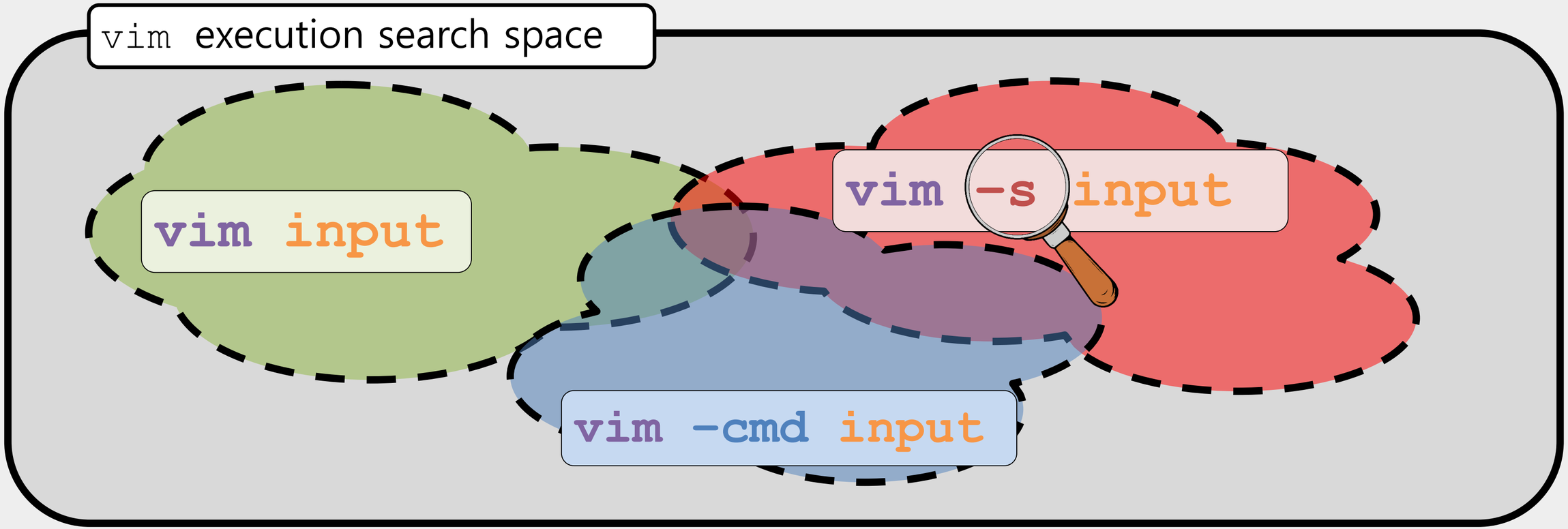
Ahcheong Lee (KAIST, South Korea)
Youngseok Choi (KAIST)
Yunho Kim (Hanyang Univ. , South Korea)
Shin Hong (Handong Global Univ., South Korea)
Moonzoo Kim (KAIST)

In Short - Mutate Command Lines as a Separated Domain!

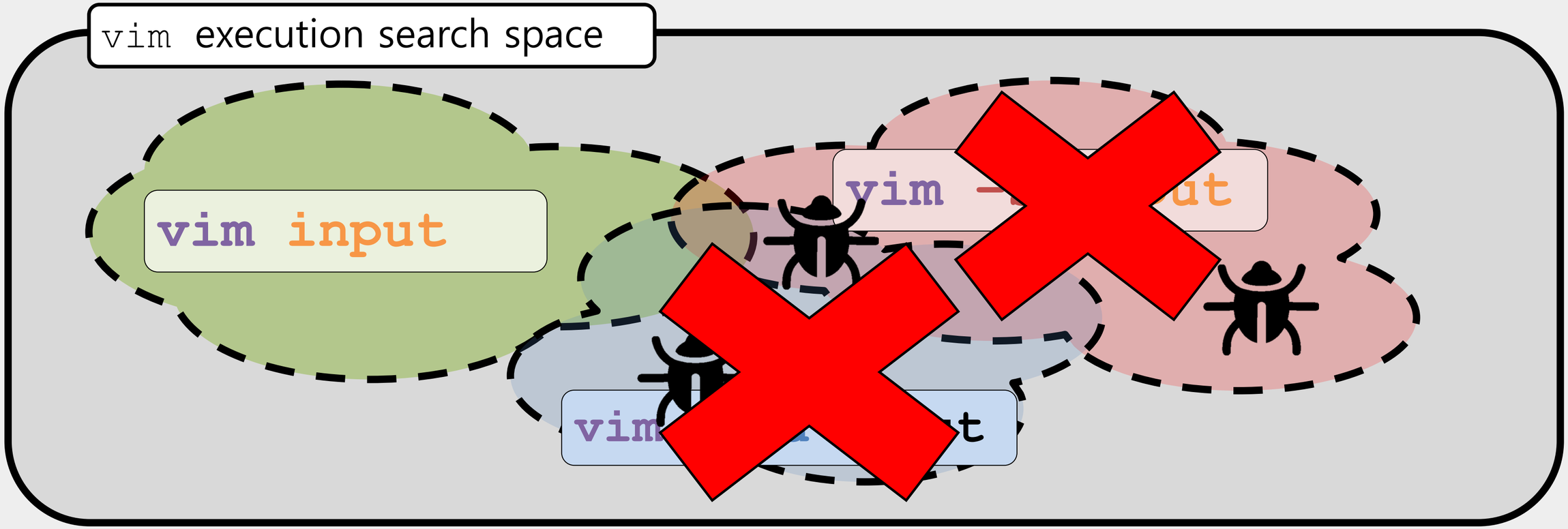
CLI (Command Line Interface) programs use **command lines**
to determine which features (functions) to execute

In Short - Mutate Command Lines as a Separated Domain!

CLI (Command Line Interface) programs use **command lines**
to determine which features (functions) to execute

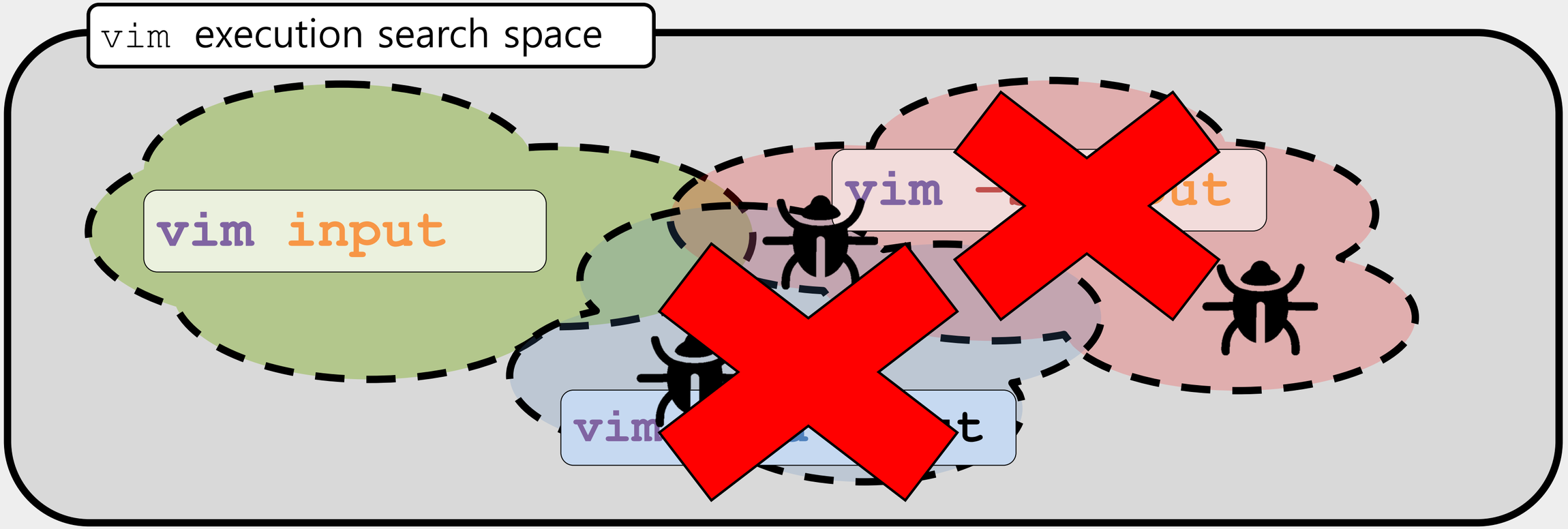


Previous Fuzzing Techniques Did Not Care Command Lines



Previous Fuzzing Techniques Did Not Care Command Lines

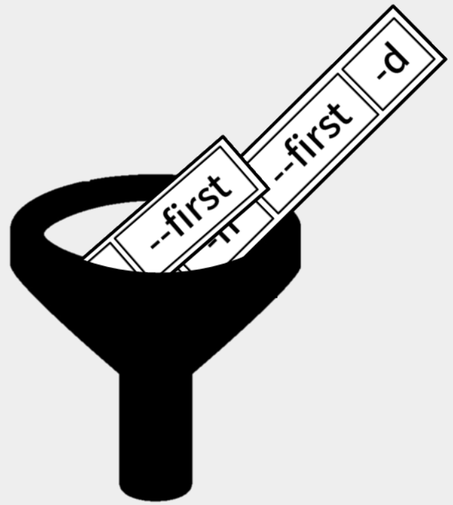
3/4 of 100 recent fuzzing papers provide NO command line information.



Contributions

Contribution 1.
POWER mutates and selects command lines in strategic way.

Contribution 2.
POWERUP found 88 crashes in 15 subjects
(which is significant more than AFL++, ConfigFuzz, Eclipser)

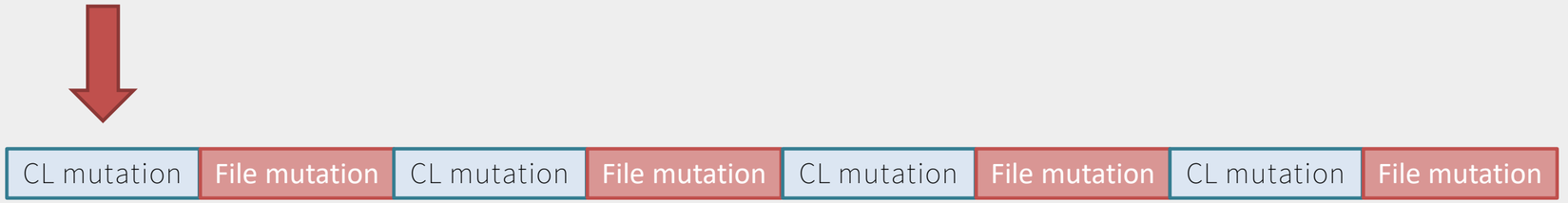


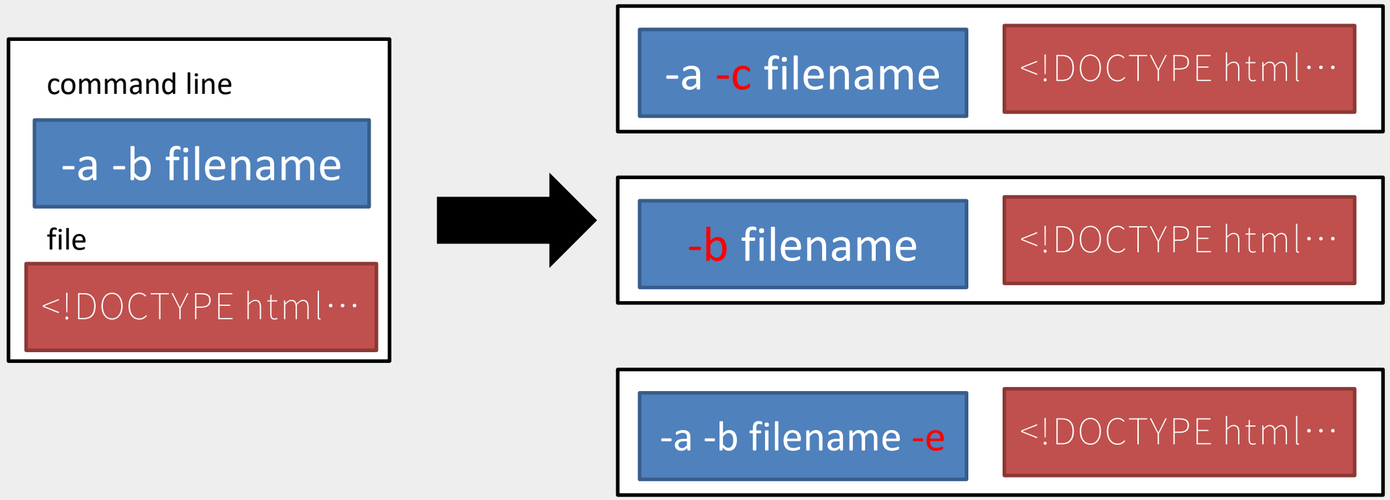
2 Key features

1. Separated mutation with interleaving
   Mutate command lines and files individually

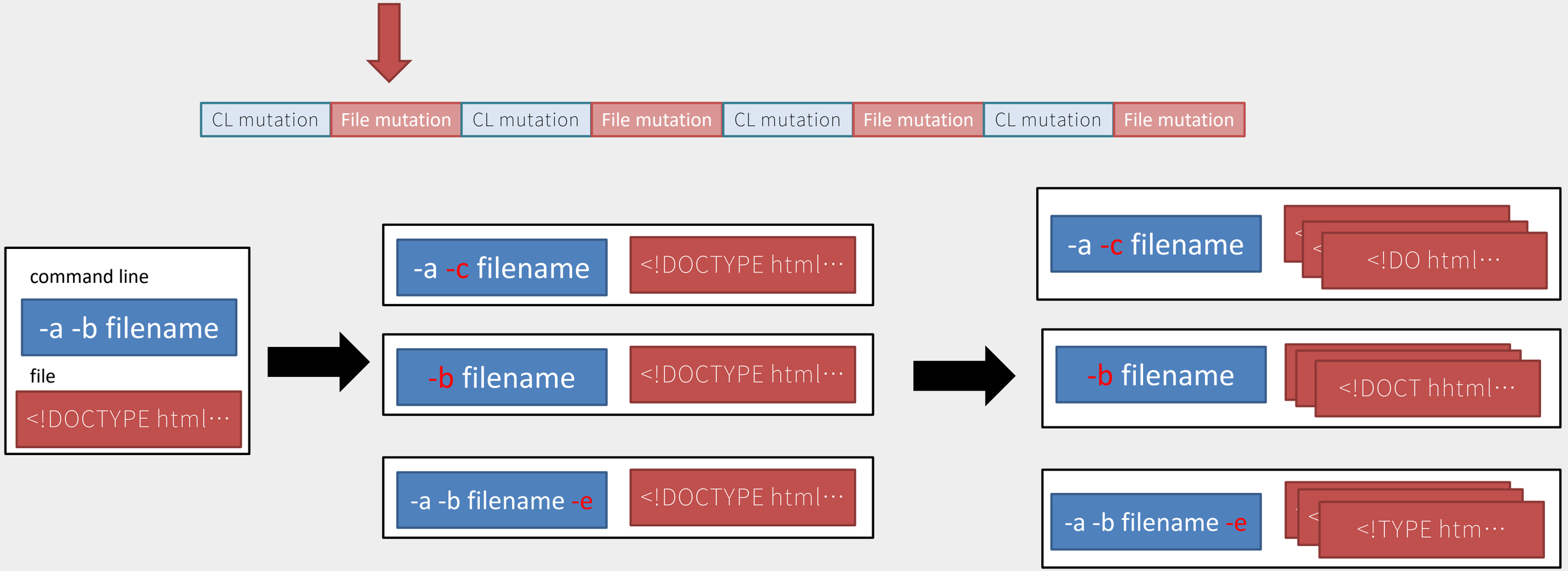
2. Distinct Command line selection
   Mutate files of selected distinct command lines only

Overall Process - Command Line Mutation

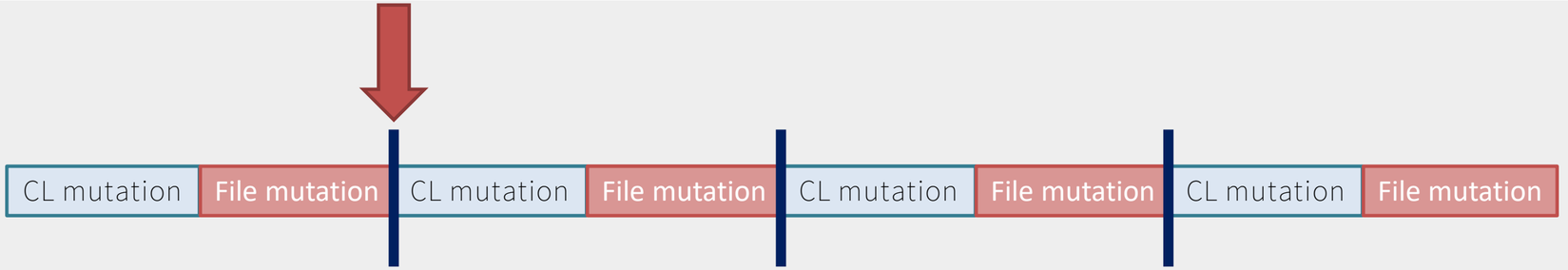


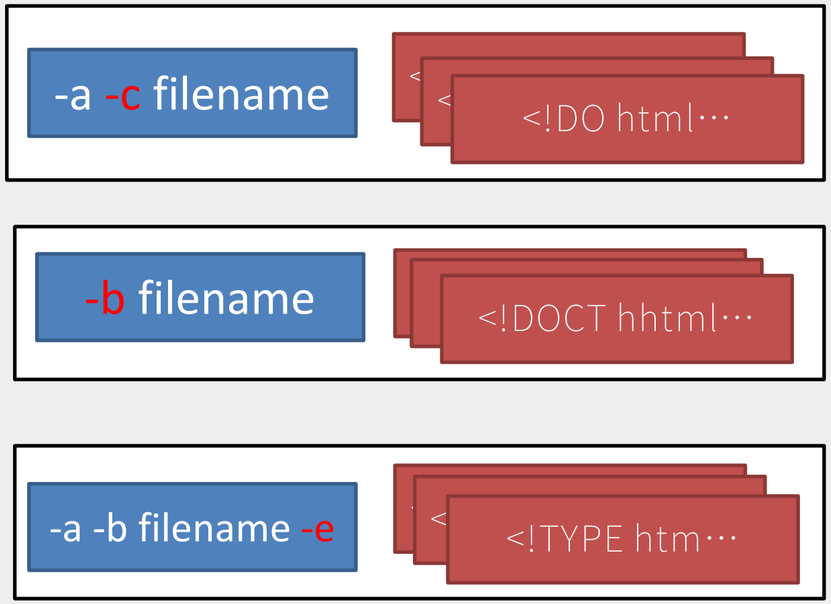


Overall Process - File Mutation



Overall Process - Corpus Shrinking (Get Distinct CL)





Strategic mutation – separated mutation

One naïve way to mutate command lines: Concatenation

1. Make a long byte string

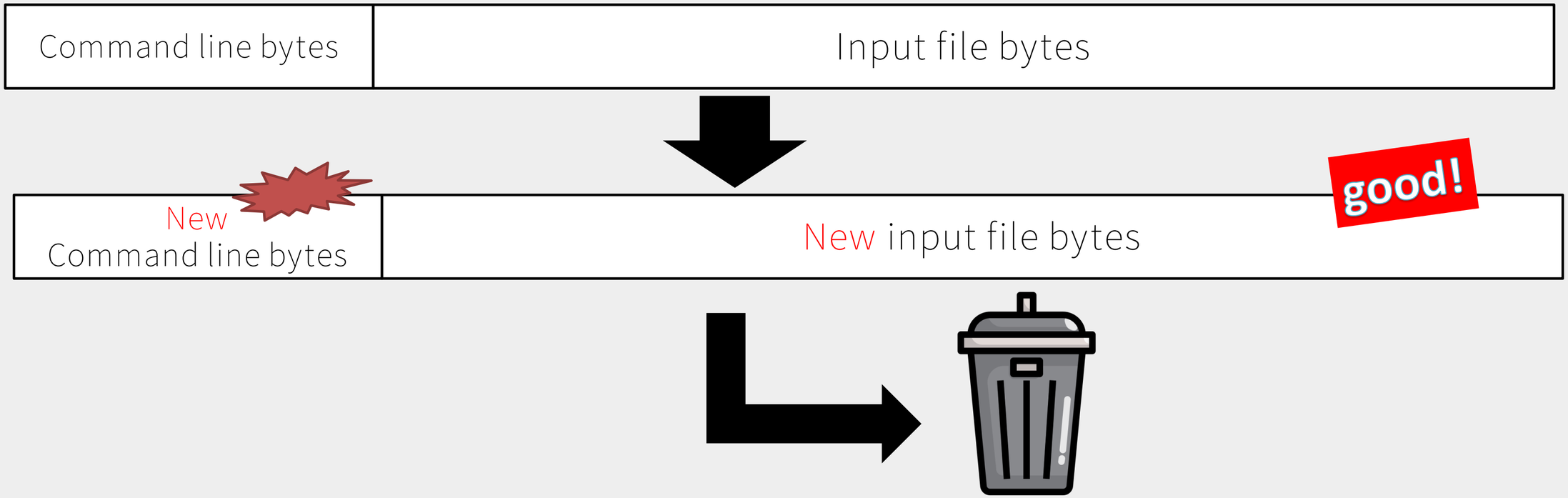
command line	input file
-a -b -c -opt -s	<!DOCTYPE html><!-- saved from url=https://www. / --><html

2. Mutation

-a -e -format -opt -s	<!DOCTYPE html><!-- saved asasdfsdfsdfs://www. / --><html

Strategic mutation – separated mutation

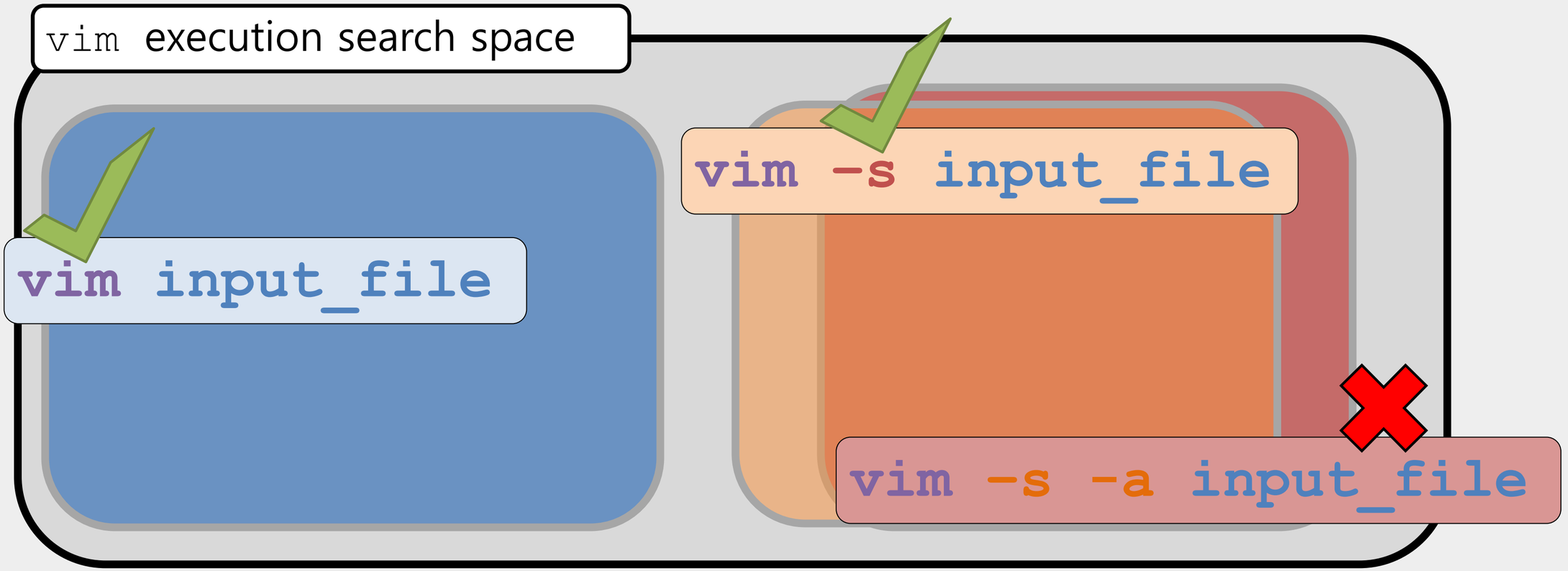
If either one of them become “invalid”, the test case become useless.



POWERUP covered 22% more branches and found 120% more crashes than the concatenation method.

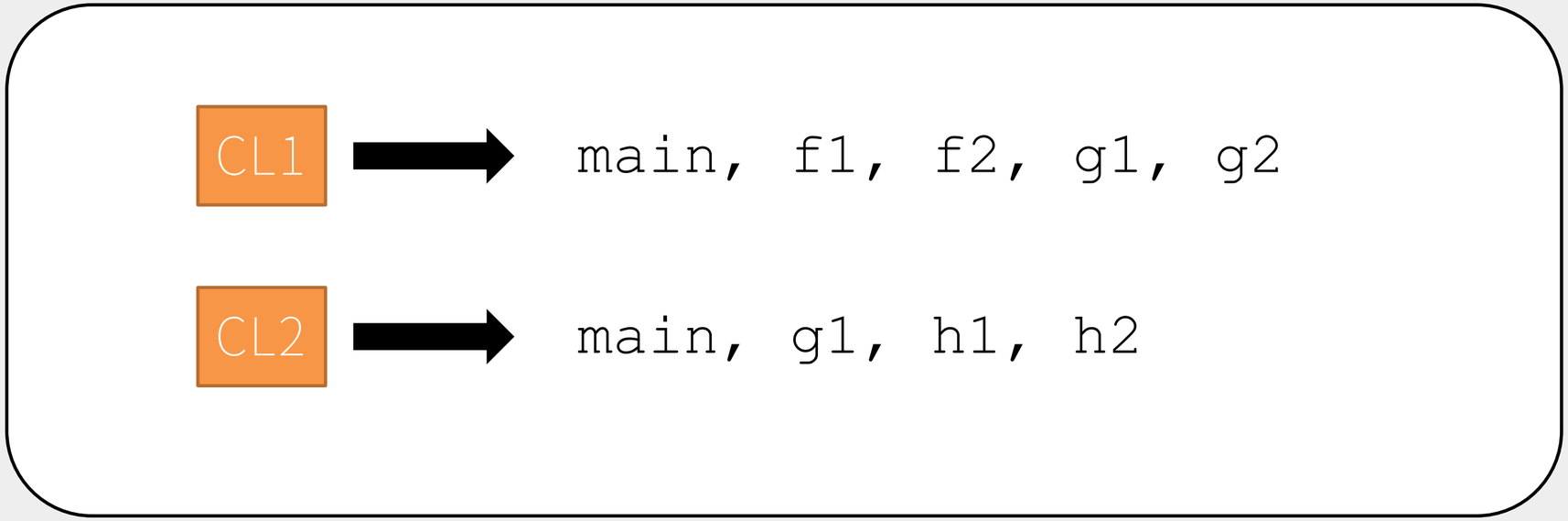
Strategic Selection of Command Lines

Focus on Distinct Command lines



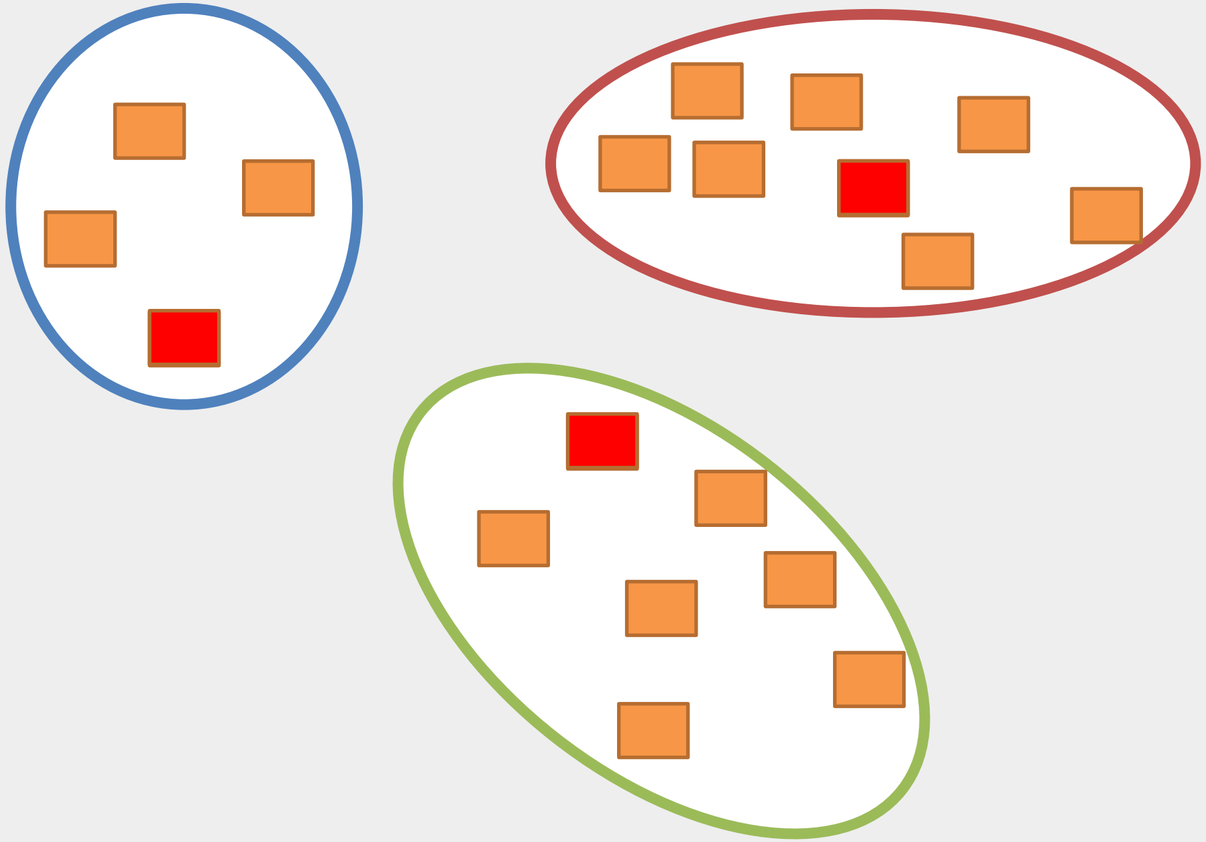
Strategic Selection of Distinct Command Lines

1. Function call profile extraction



Strategic Selection of Distinct Command Lines

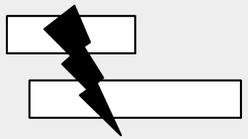
2. Clustering (K-means)



$$\text{Distance} = 1 - \frac{|A \cap B|}{|A \cup B|}$$

Evaluation

RQ1 Are POWER and POWERUP better than other command-line fuzzing techniques like Configfuzz [TOSEM], AFL++-argv, and Eclipser?



RQ2 Did separated mutation contribute to POWERUP's performance?

Measurement: the # of detected unique crashes

Setup : 12 hours run, repeated 5 times.

Test subjects

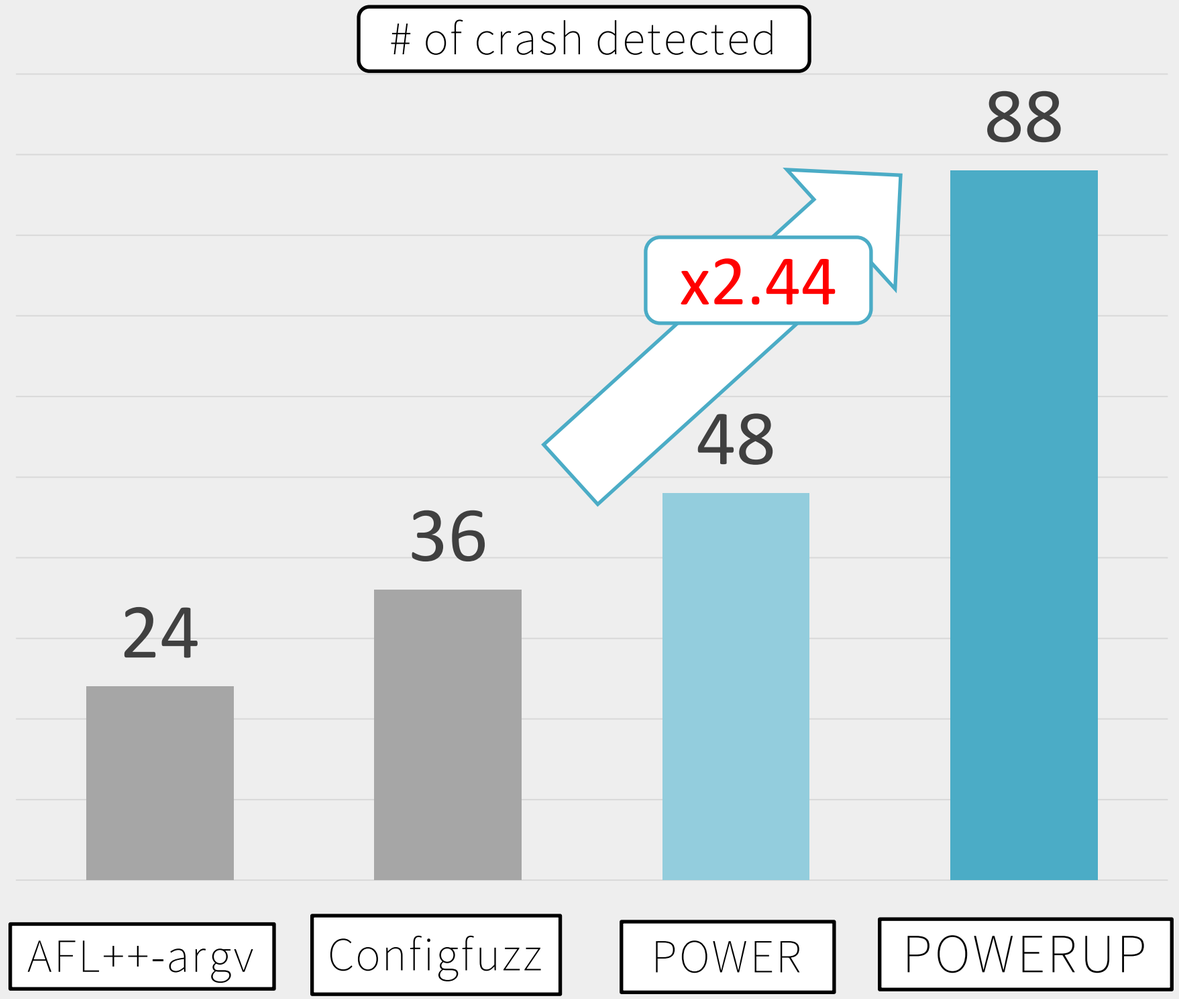
20 C/C++ open source programs that are used by other fuzzing papers

Subjects	Size (Loc)	# of CLI keywords	Subjects	Size (Loc)	# of CLI keywords
avconv	454,936	761	nasm	70,903	218
bison	54,423	51	objdump	877,165	84
cjpeg	6,308	33	pdftohtml	38,111	26
dwarfdump	83,545	103	pdftopng	97,890	18
exiv2	33,417	76	pspp	4,901	20
ffmpeg	774,186	1817	readelf	74,789	98
gm	197,891	757	tiff2pdf	8,234	30
gs	1,174,673	350	tiff2ps	5,646	34
jasper	2,920	20	xmllint	11,285	66
mpg123	11,298	122	xmlwf	4,147	15

Avg. Loc : 199,333

Avg. # of keywords : 235
Median # of keywords : 71

Result – RQ 1 : Are POWER and POWERUP better than other command-line fuzzing techniques?



AFL++-argv:
AFL++'s feature that mutates only command lines

ConfigFuzz:
A tool that can generate test driver from json grammar file, so that a fuzzer can mutate both command lines and input files

Result – Crash example

One crash found in `mpg123` contains 13 command line keywords:

```
./mpg123 -smooth -listentry -z -w 1 --quiet --index - -4to1
-2 -q -fifo --outfile
```

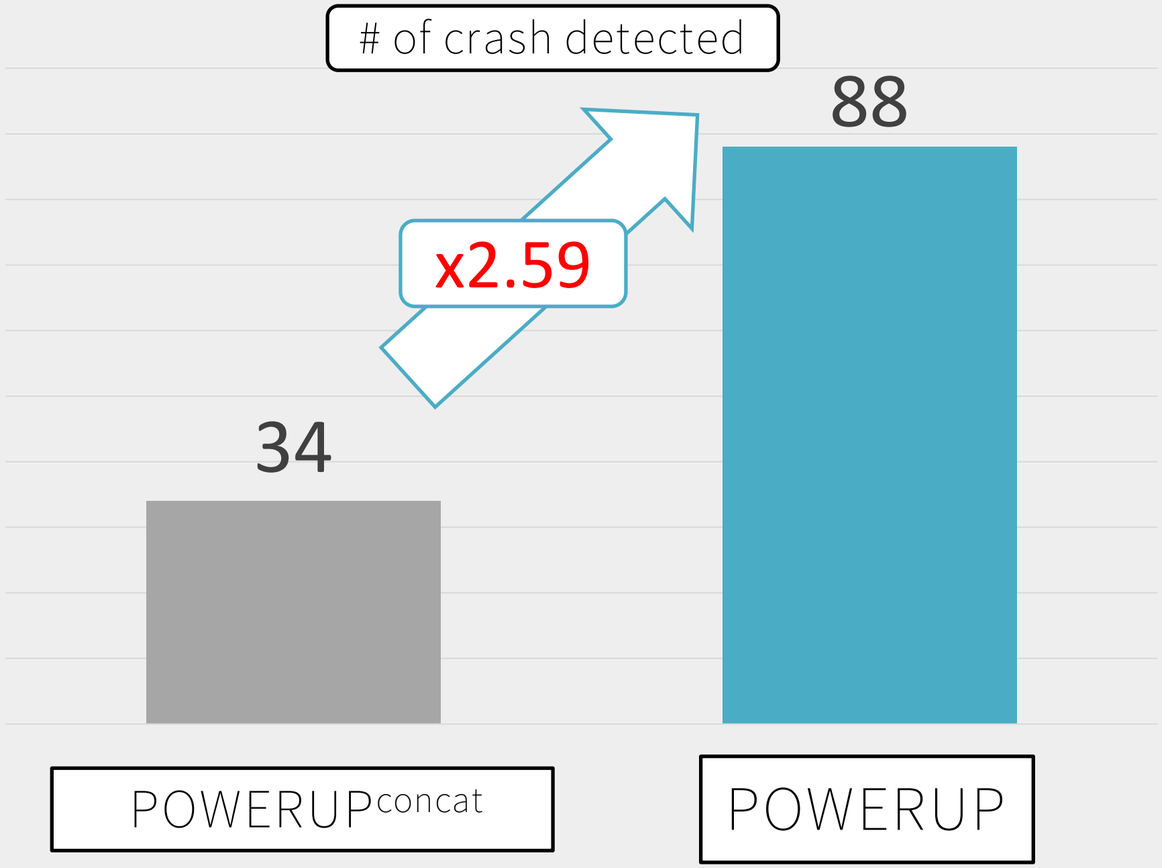
The developer of `mpg123` commented that:

> “Interesting approach you find stuff where oss-fuzz didn’t anymore.”

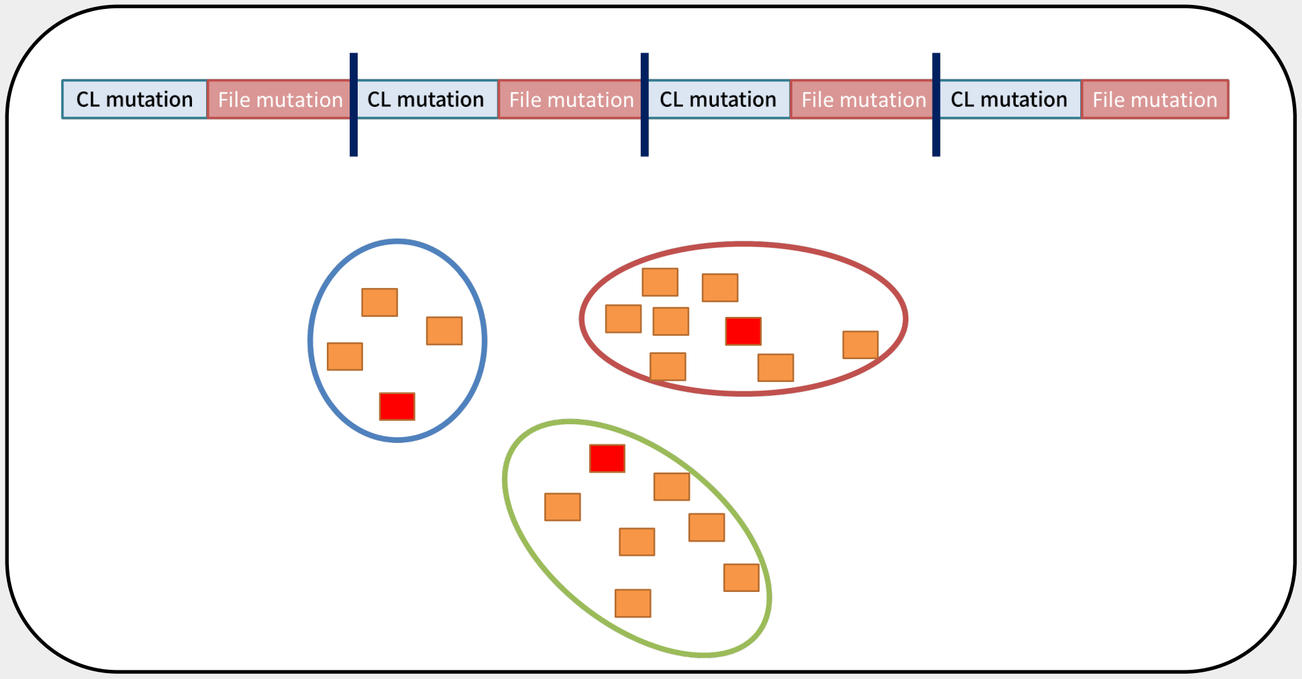
Result – RQ2

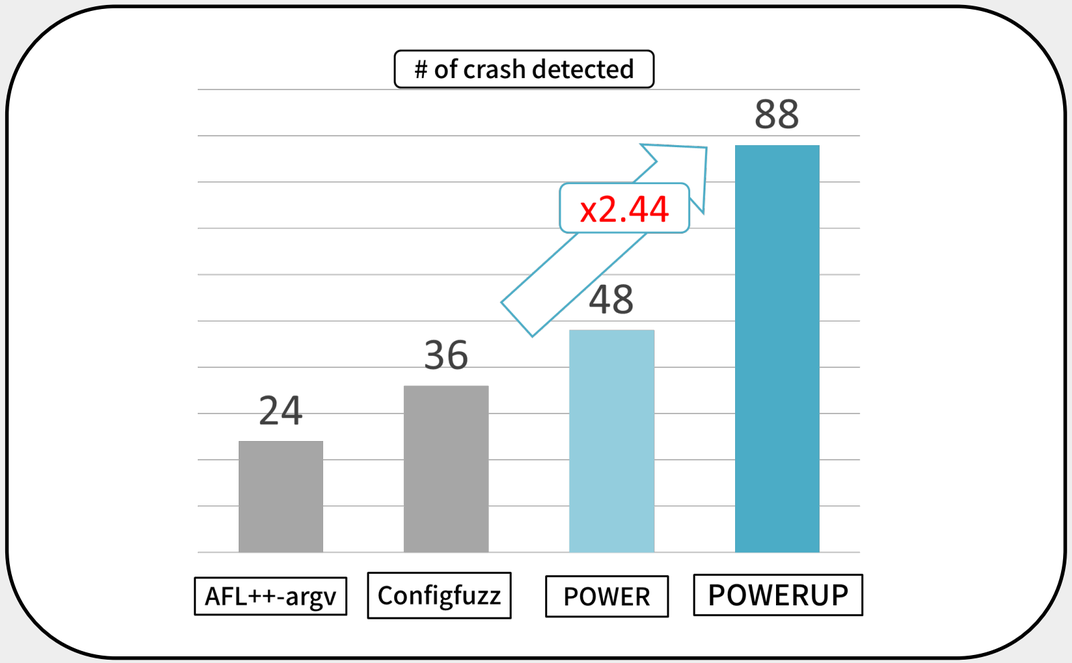
RQ2: Did separated mutation contribute to POWERUP's performance?

- POWERUP^{concat}



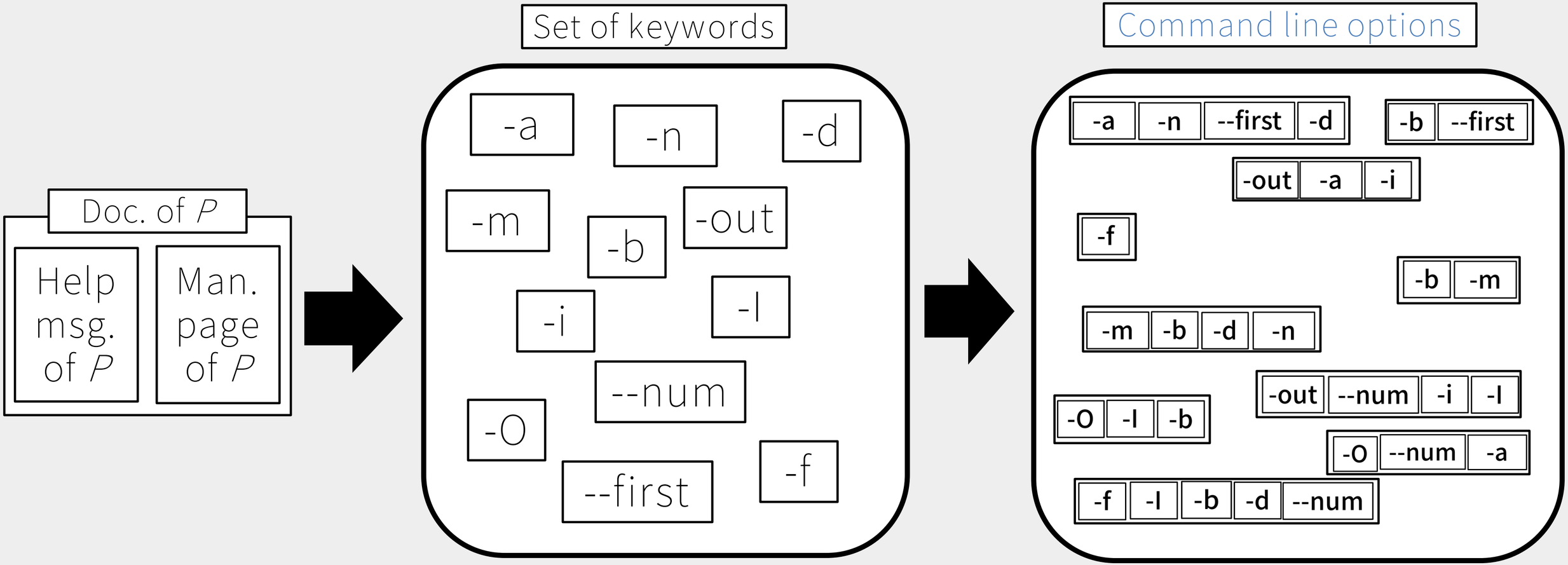
Conclusion


CL mutation
File mutation
CL mutation
File mutation
CL mutation
File mutation
CL mutation
File mutation


# of crash detected
88
x2.44
48
36
24
AFL++-argv
Configfuzz
POWER
POWERUP

Careful Generation of Command Line Options

Dictionary-based mutation to generate valid command line options



Careful Selection of Distinct Command Line Options

Generated Command line options

-a -n --first -d

-b --first

-out -a -i

-f

-b -m

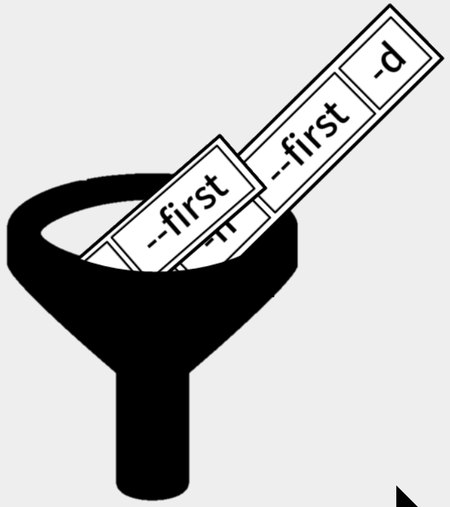
-m -b -d -n

-out --num -i -l

-O -l -b

-O --num -a

-f -l -b -d --num



Selection

command line options that are far different from each other

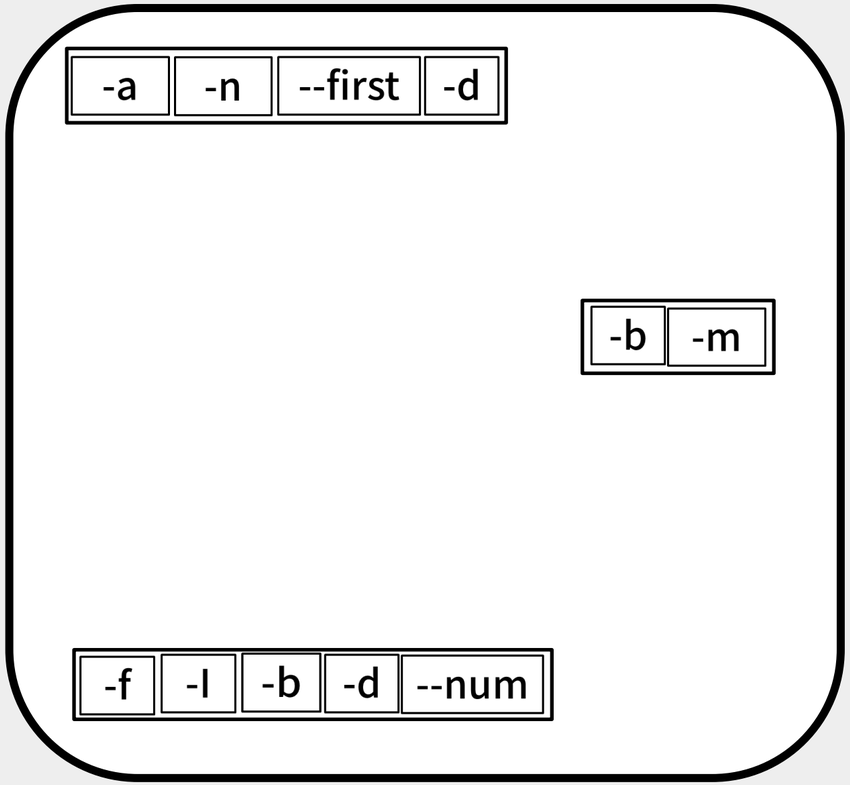
-a -n --first -d

-b -m

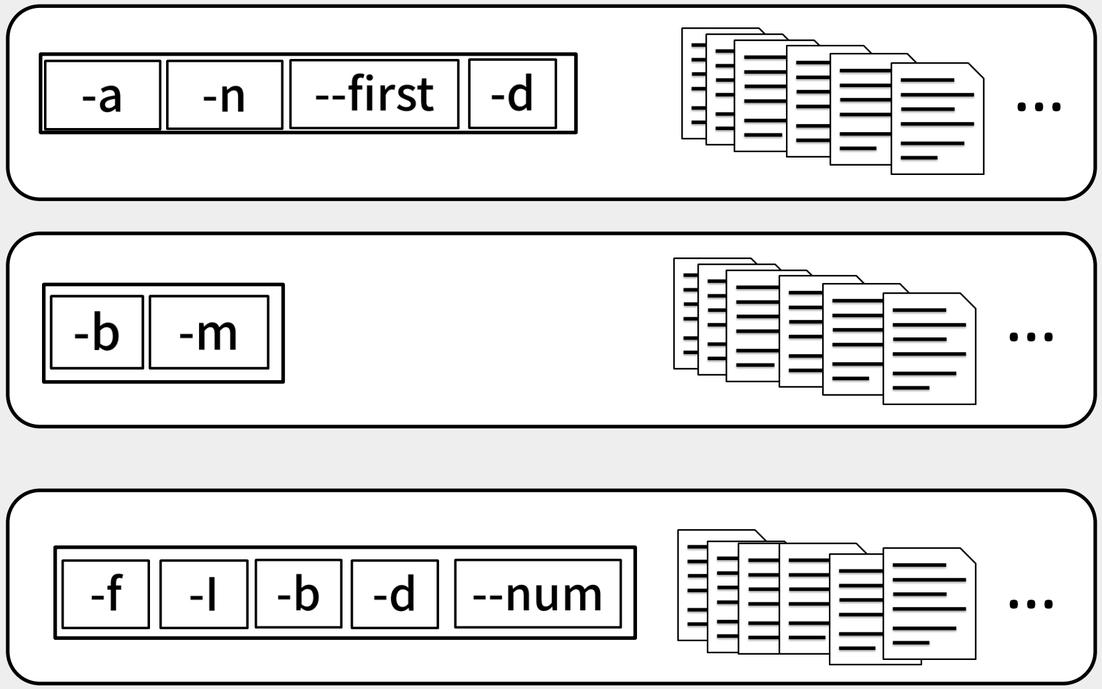
-f -l -b -d --num

Careful Selection of Distinct Command Line Options

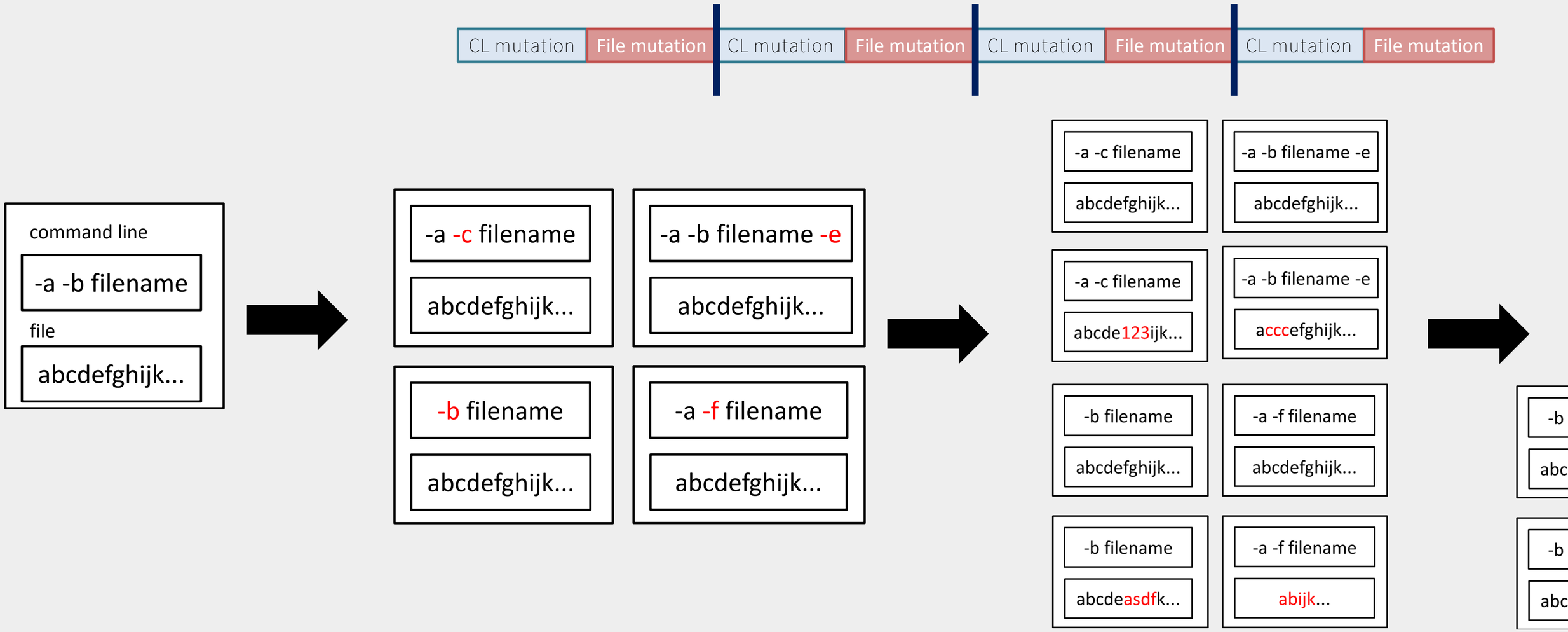
command line options
that are far different from each other



Mutate input file
with selected command line options



Overall Process - Command Line Mutation



Strategic mutation – separated mutation

A program is a mixture of command line constraints and input file constraints

```
1   process_args(argc, argv, &gf);   // Command check, set global flags
2   filename = argv[argc - 1];
3
4   open_and_read_type(filename, &filetype); // File checks (magic bytes, …)
5
6   if ((filetype == 2) && (gf.name == "ELF"))) {
7     process_one_file(filename, …);  // Main logic
8   } else {
9     exit(EXIT_FAILURE);
10  }
```

POWERUP covered 22% more branches and found 120% more crashes than the concatenation method.