CarpetFuzz: Automatic Program Option Constraint Extraction from Documentation for Fuzzing

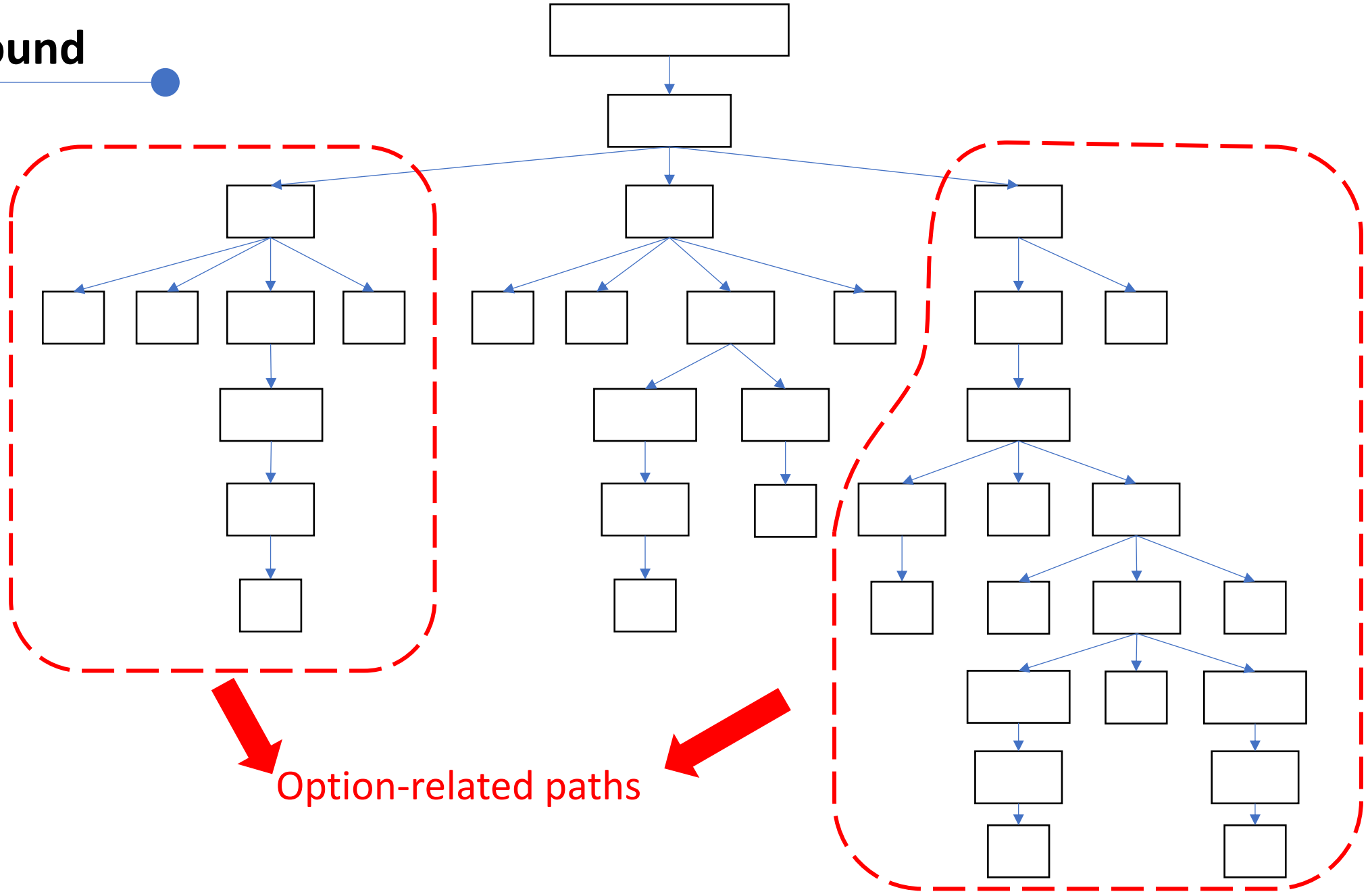
Dawei Wang, Ying Li, Zhiyu Zhang,

Kai Chen

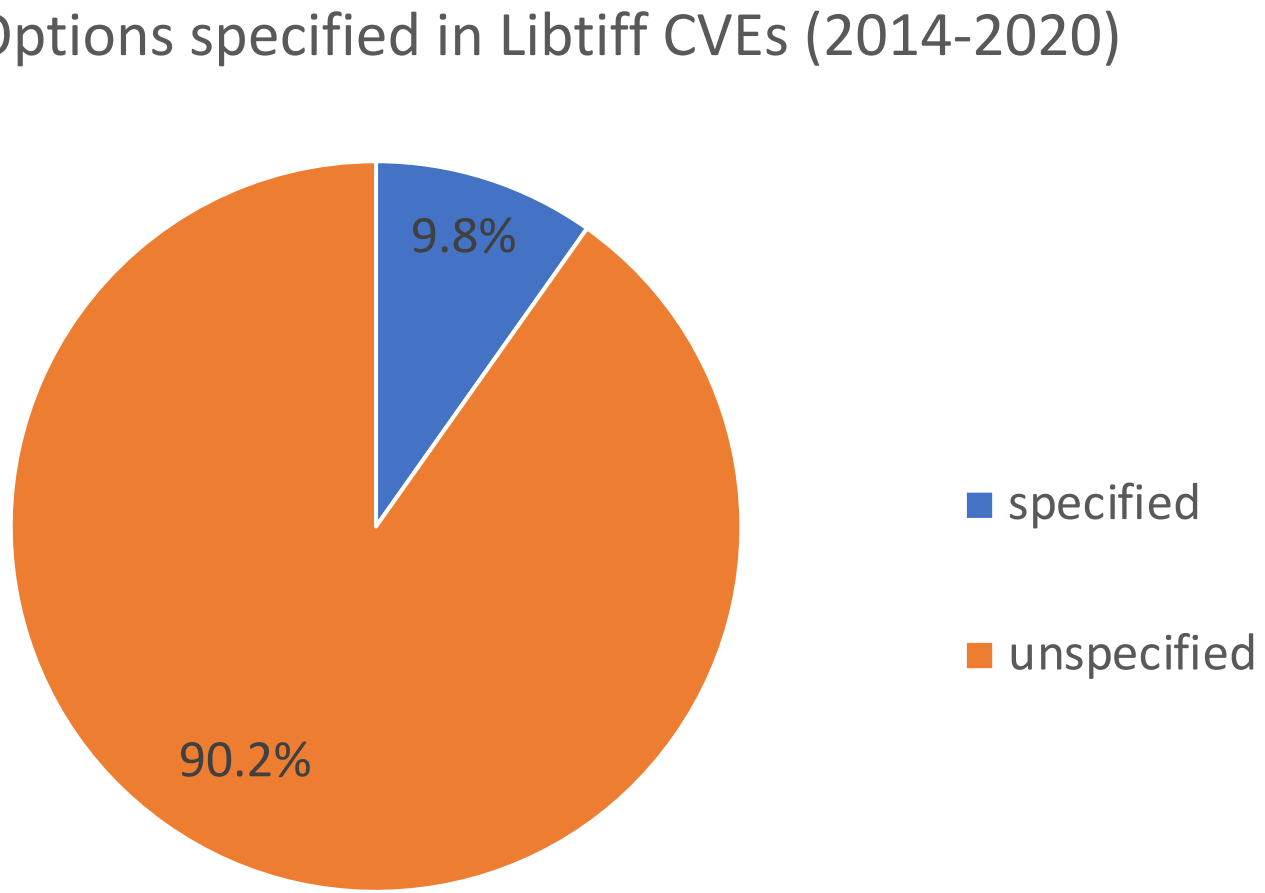
中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

USENIX Security 2023

Background



Background

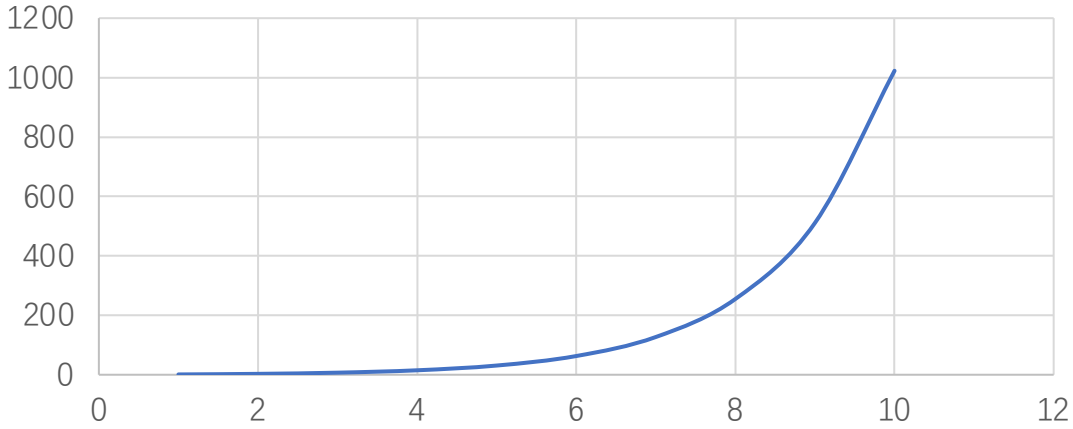


Many option-related paths may remain unexplored

Background

ImageMagick

242 options



7.1×10^{72} combinations

Impossible to traverse all combinations

Background

pdftops manpage

-form

Generate a PostScript form which can be imported by software that understands forms. A form contains a single page, so if you use this option with a multi-page PDF file, you must use -f and -l to specify a single page. The -level1 option cannot be used with -form. No more than one of the mode options (-origpagesizes, -eps, -form) may be given.

tiffcrop manpage

-O portrait|landscape|auto

Set the output orientation of the pages or sections. Auto will use the arrangement that requires the fewest pages. This option is only meaningful in conjunction with the -P option to format an image to fit on a specific paper size.

lrzip manpage

-O

Set the output directory for the default filename. This option cannot be combined with -o.

when not satisfied

```
# pdftops –level1 –form in.pdf /tmp/foo
Error: forms are only available with Level 2 output
```

Challenge

Various declaration ways →

- Not meaningful in conjunction with
- Cannot be combined with
- Cannot be used with

Implicit declared relationships →

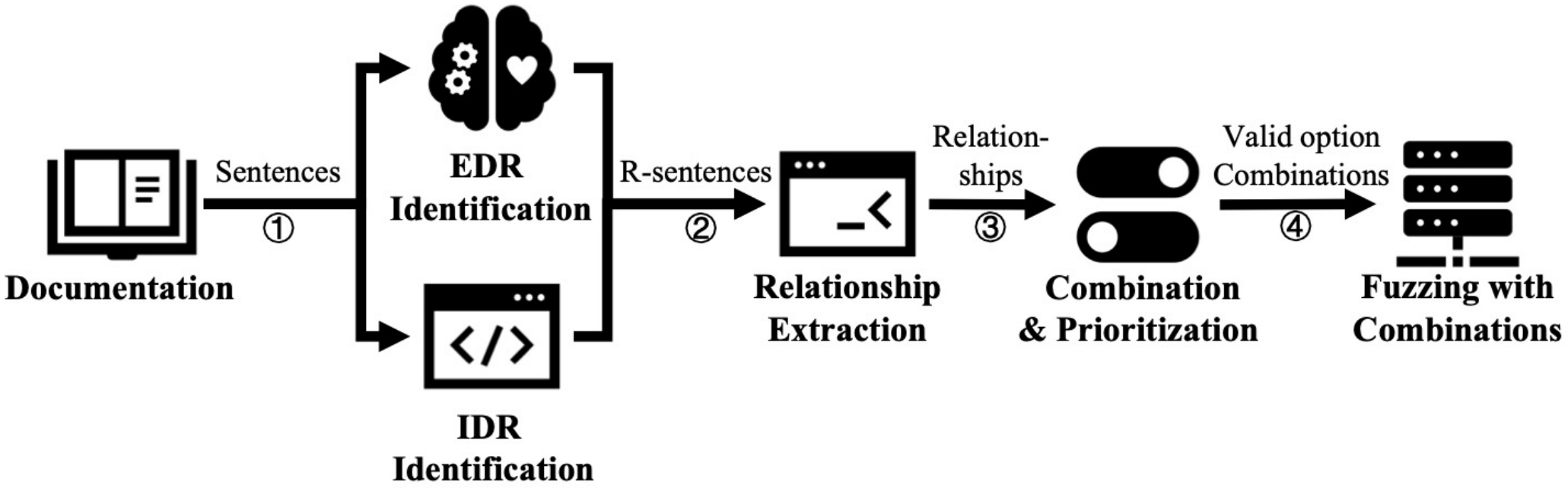
- –B: Force output to be written with Big-Endian byte order.
- –L: Force output to be written with Little-Endian byte order.

Hard to extract concrete relationships →

- Either -f or -b must be used with -C, and -C cannot be used with -F or –d.

Approach

Overview



Approach

☐ Explicit Declared Relationship (EDR) Identification

- 37,672 programs
- **302,875** sentences

- **1,817** labeled sentences
- Only **0.6%**

Approach

☐ Implicit Declared Relationship (IDR) Identification

- Approach: Check predicate, object, and grammatical structure

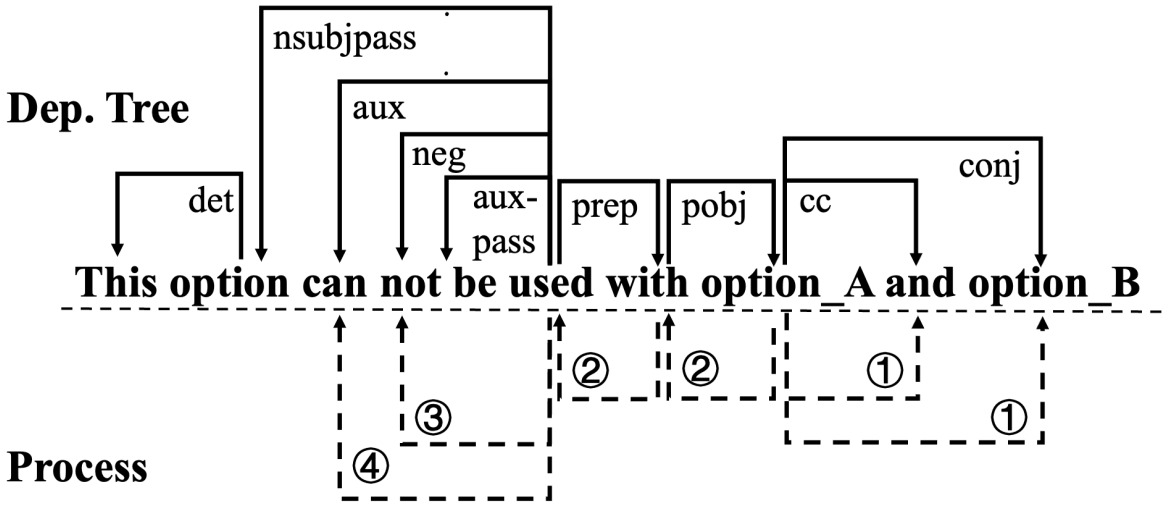
-L: Force output to be written in Little-Endian byte order.
-B: Force output to be written in Big-Endian byte order.

- Same/antonyms predicate
- Same object
- Parallel grammatical structure

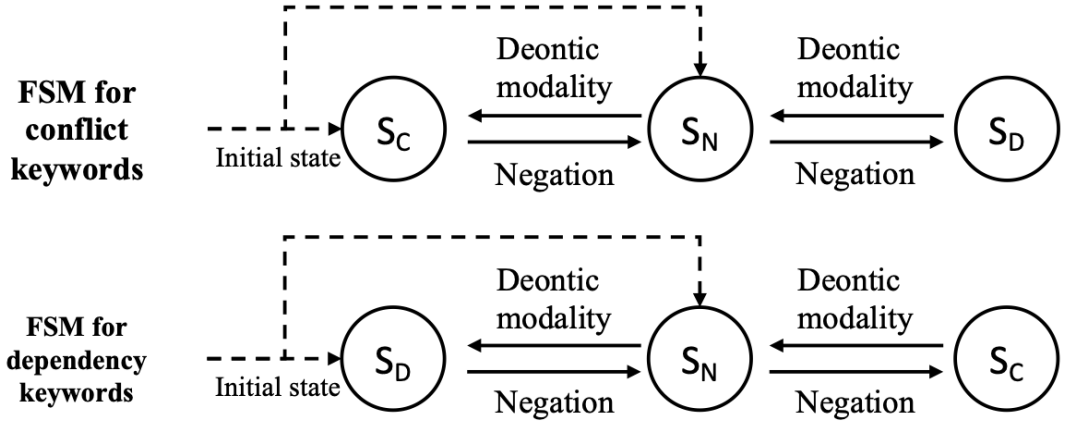
These options do the **same (or opposite) things** on the **same object**

Approach

☐ Relationship Extraction



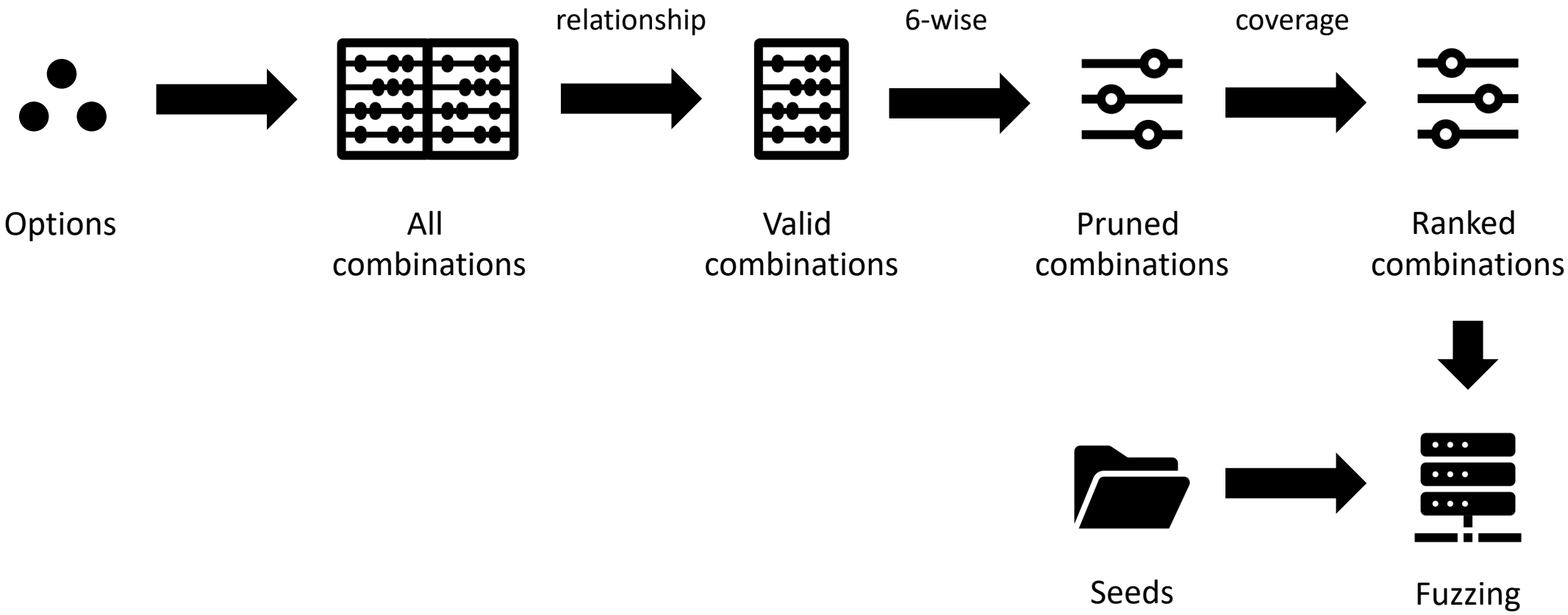
A mixing forward and backward traversal method



A polarity-based finite state machine

Approach

☐ Combination, Prioritization, and Fuzzing



Evaluation

20 programs

- Formats: 11

Program	Package	Format	#OPT.	#REL.
cmark	cmark-git-9c8e8	md	9	3
editcap	wireshark-4.0.1	pcap	31	4
eu-elfclassify	elfutils-0.188	elf	22	71
img2sixel	libsixel-git-6a5be	jpg	28	6
jpegoptim	jpegoptim-1.5.0	jpg	34	29
jpegtran	libjpeg-turbo-2.1.4	jpg	21	1
jq	jq-1.6	json	24	4
lrzip	lrzip-0.651	lrz	22	10
ogg123	vorbis-tools-1.4.2	ogg	13	1
openssl-asn1parse	openSSL-git-31ff3	pem	12	2
openssl-ec	openSSL-git-31ff3	pem	16	3
openssl-rsa	openSSL-git-31ff3	pem	28	71
pdftops	xpdf-4.0.3	pdf	30	30
pdftotext	xpdf-4.0.3	pdf	24	5
podofoencrypt	podofo-0.9.8	pdf	11	1
speexdec	speex-1.2.1	spx	12	18
tcpprep	tcpreplay-4.4.2	pcap	19	17
tcpreplay	tcpreplay-4.4.2	pcap	29	20
tiffcp	libtiff-git-b51bb	tiff	18	4
tiffcrop	libtiff-git-b51bb	tiff	33	6

Evaluation

20 programs

Extract → **282** relationships

- Formats: 11

Filter → **67.91%** combinations

- Precision: **96.10%**
- Recall: **88.85%**

Increase → **45.97%** unique paths

Find → **57** crashes

- 0-day: **43**
- CVE: **30**

CVE-2022-4450 (OpenSSL)

Found by CarpetFuzz. Found by Dawei Wang. Found by Marc Schönefeld. Fix developed by Kurt Roeckx. Fix developed by Matt Caswell.

- Fixed in OpenSSL 3.0.8 (git commit) (Affected since 3.0.0)
- Fixed in OpenSSL 1.1.1t (git commit) (Affected since 1.1.1)

Thank You
For You Attention

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

Dawei Wang: wangdw.augustus@gmail.com
Kai Chen: chenkai@iie.ac.cn