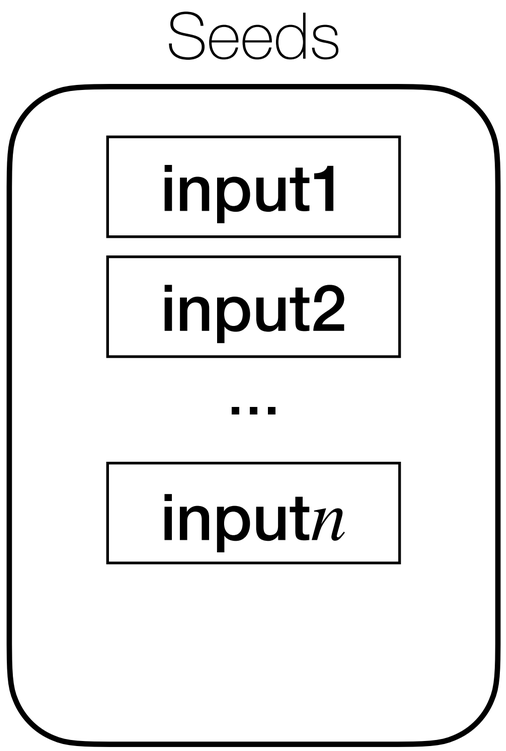
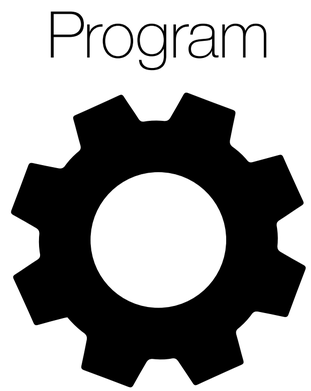
NEZHA:
Efficient Domain-Independent Differential Testing

Theofilos Petsios, Adrian Tang, Salvatore Stolfo, Angelos D. Keromytis, and Suman Jana (S&P '17)

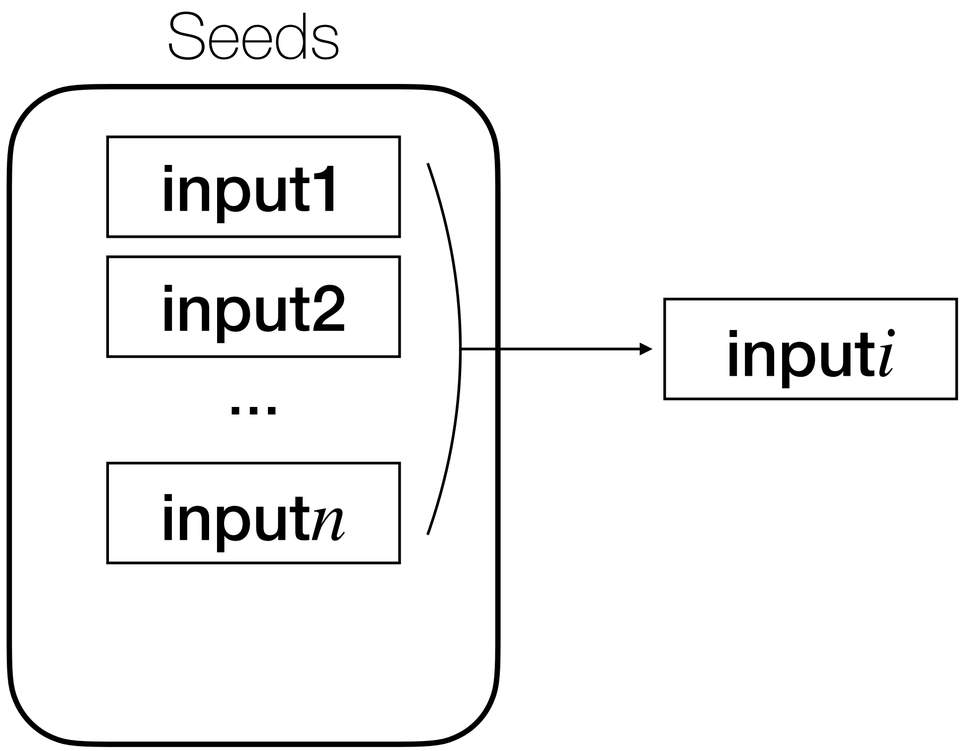
Jaemin Hong 20203698

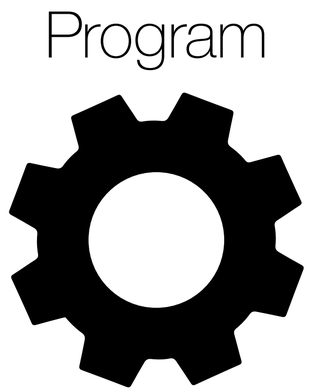
A fuzzer runs a program with random inputs.



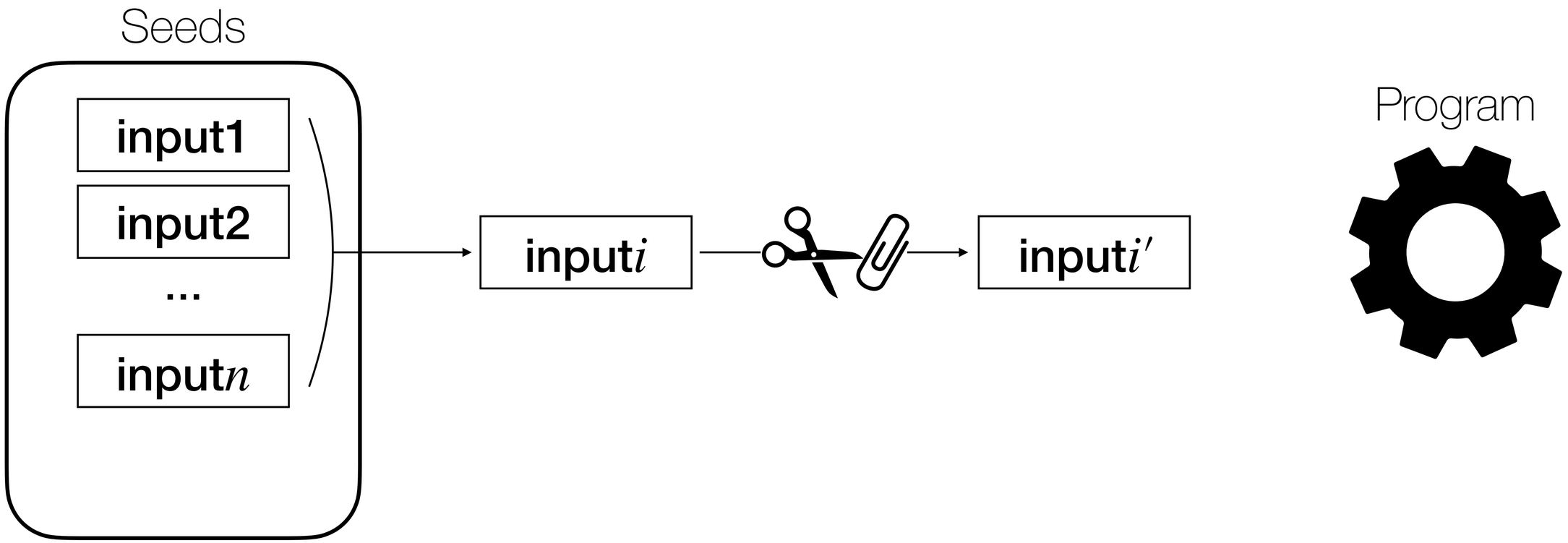


A fuzzer runs a program with random inputs.

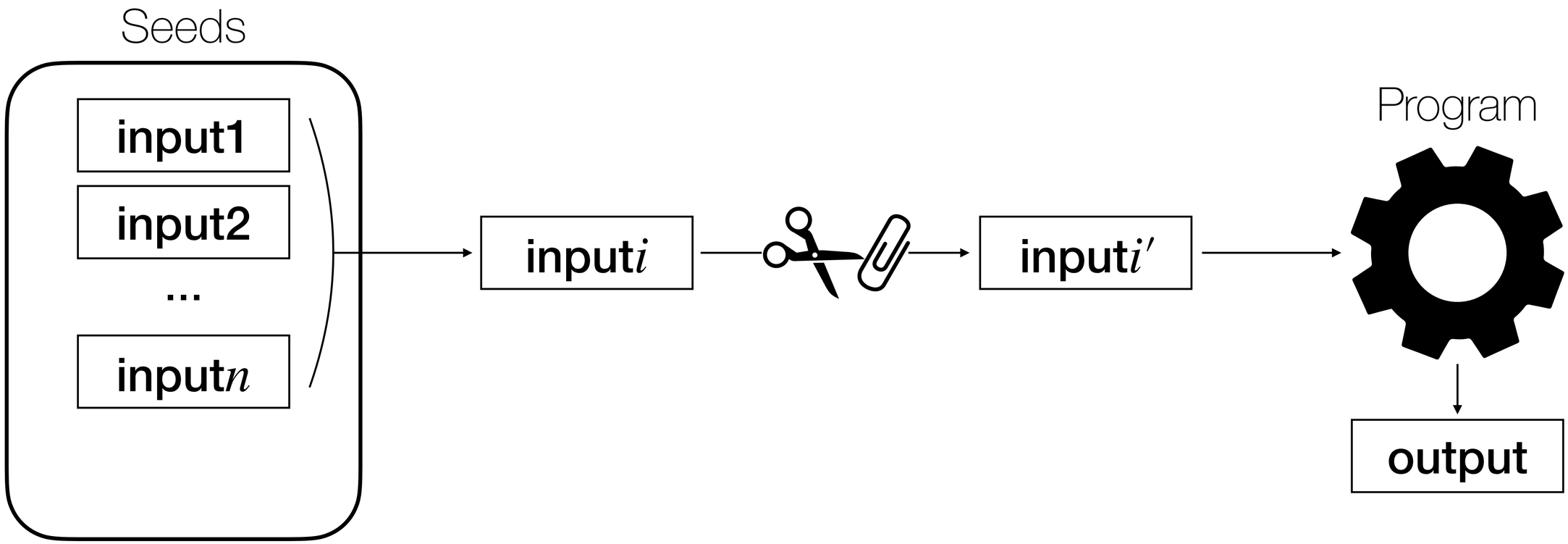




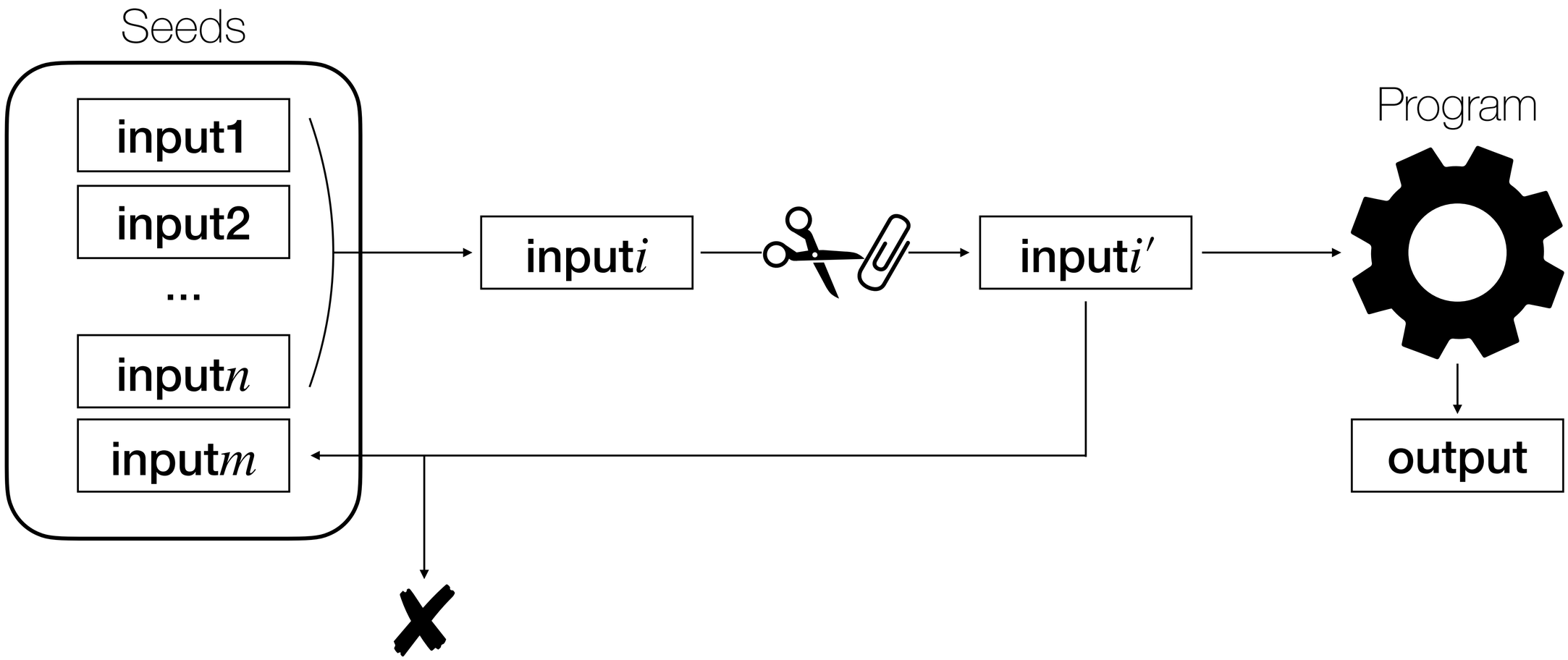
A fuzzer runs a program with random inputs.



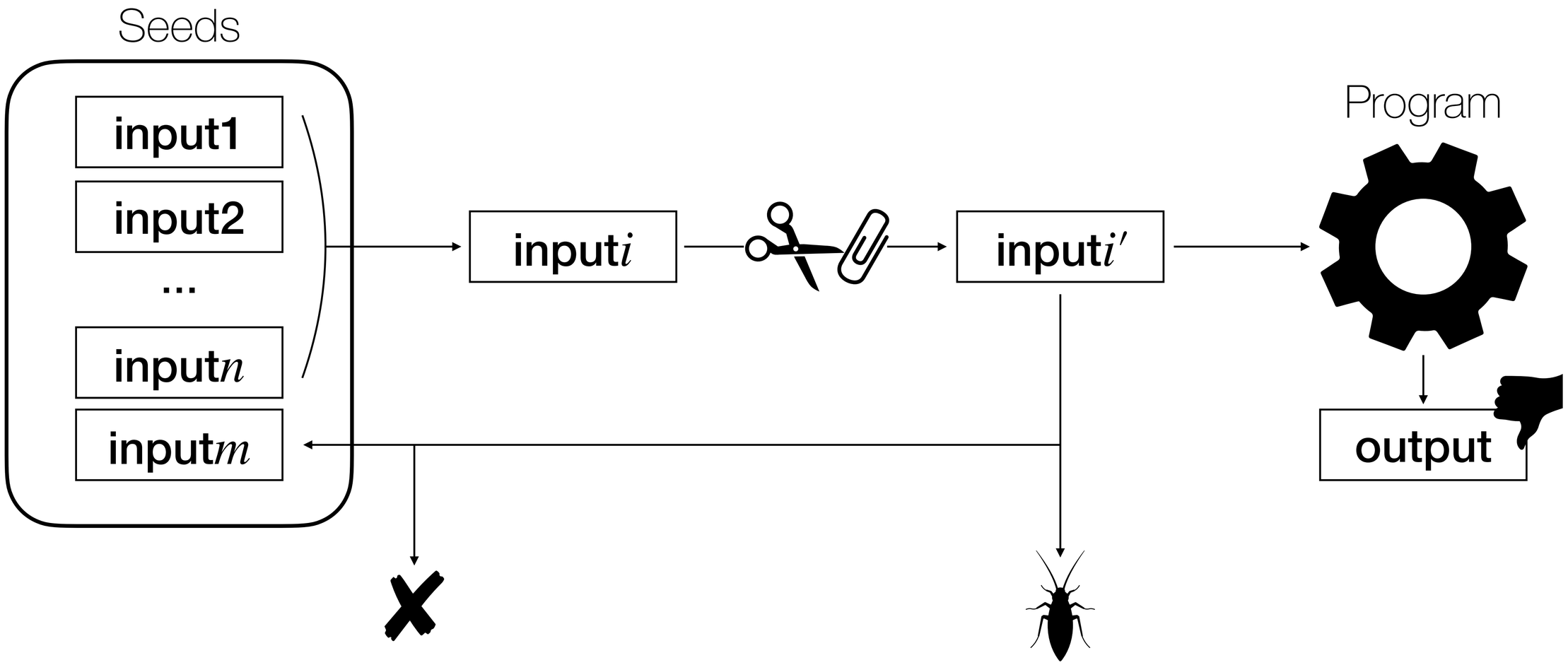
A fuzzer runs a program with random inputs.



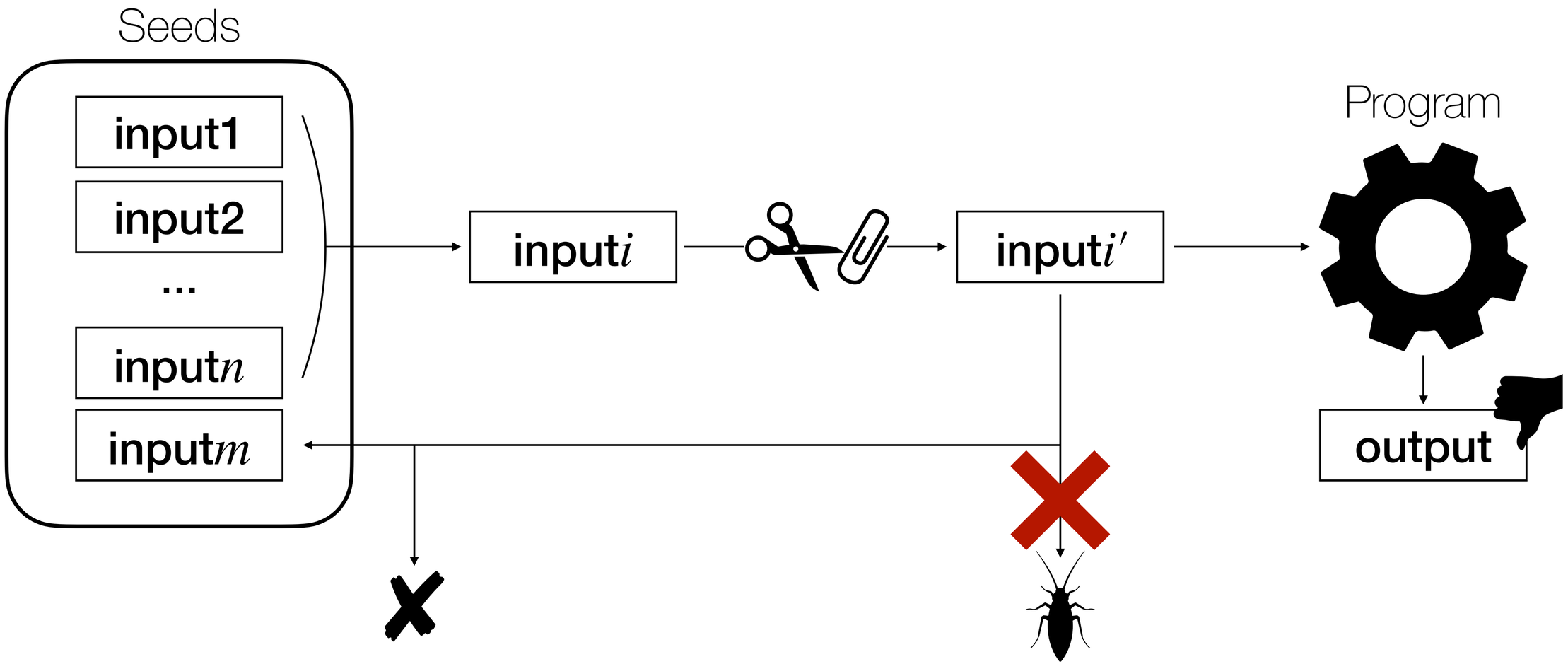
A fuzzer runs a program with random inputs.



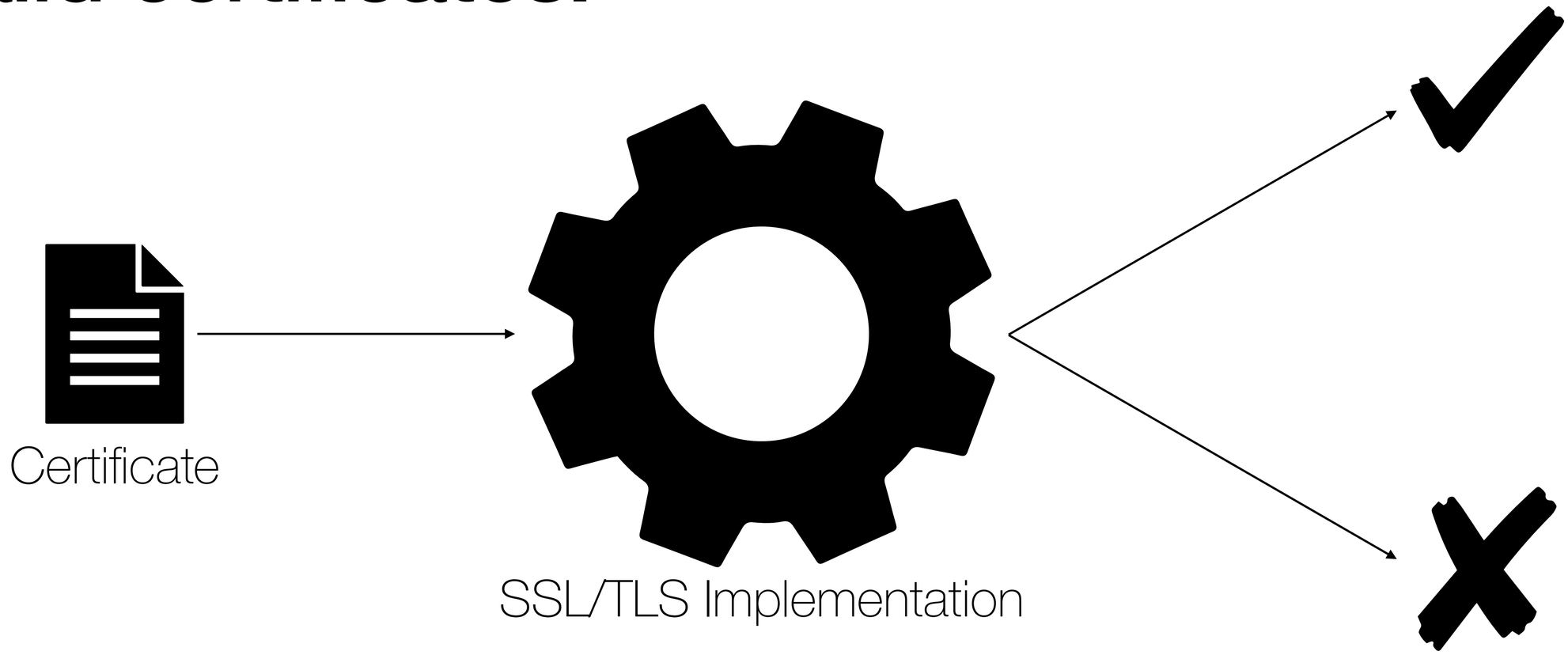
A fuzzer runs a program with random inputs.



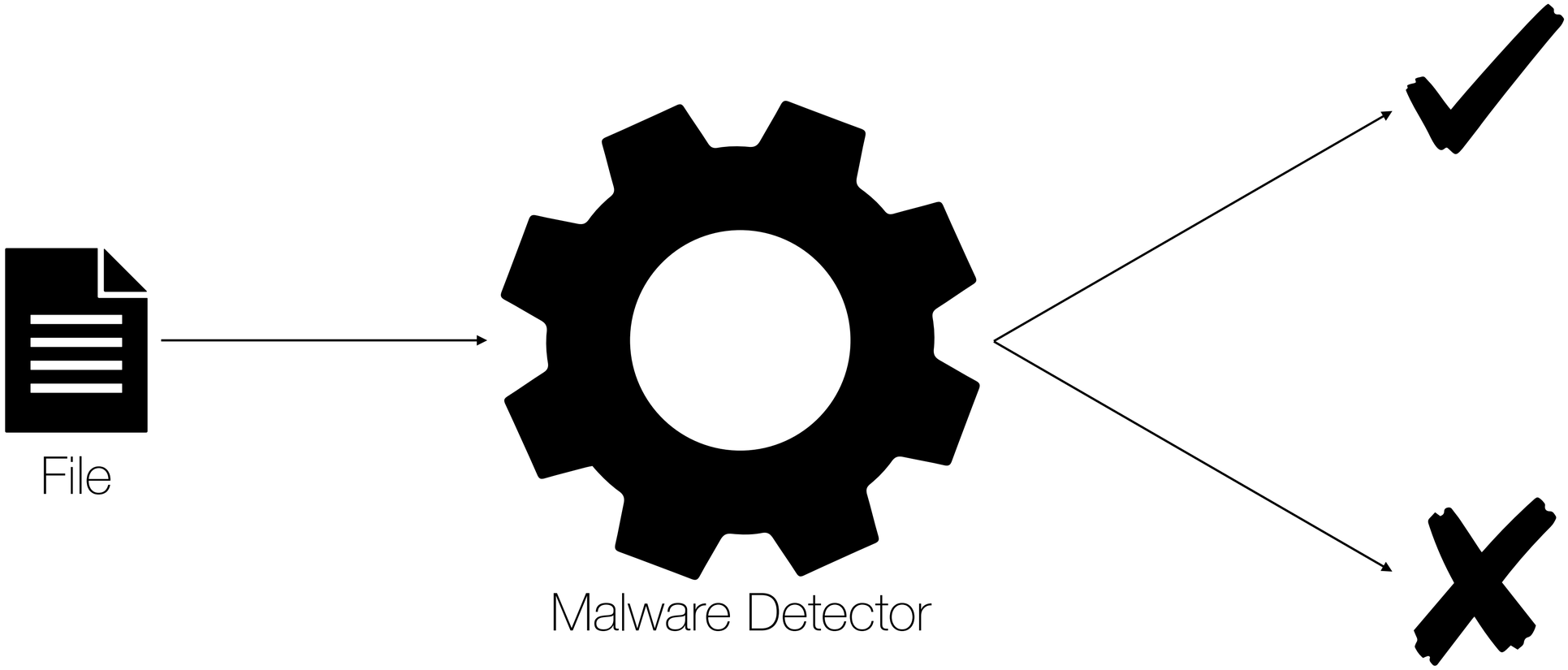
Most semantic bugs are silent.



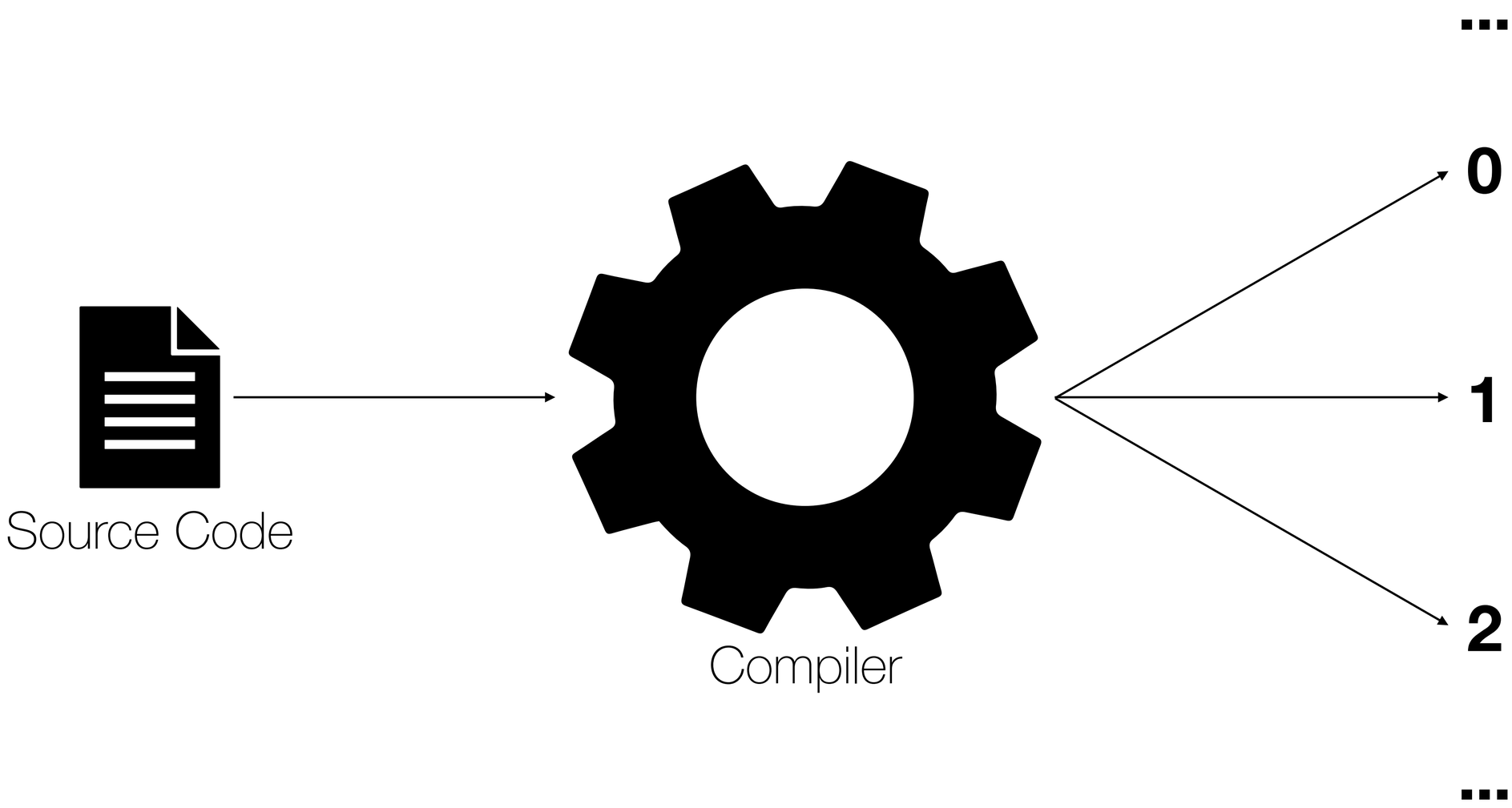
SSL/TLS implementations may accepts invalid certificates.



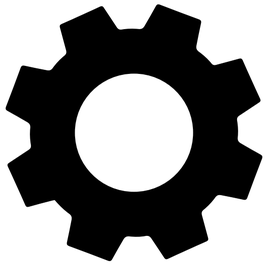
Malware detectors may miss malicious files.



Compilers may produce wrong binaries.

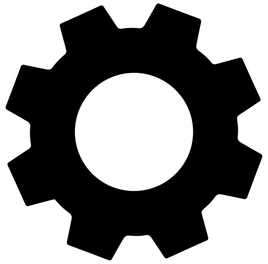


Various SSL/TLS implementations exist.



LibreSSL

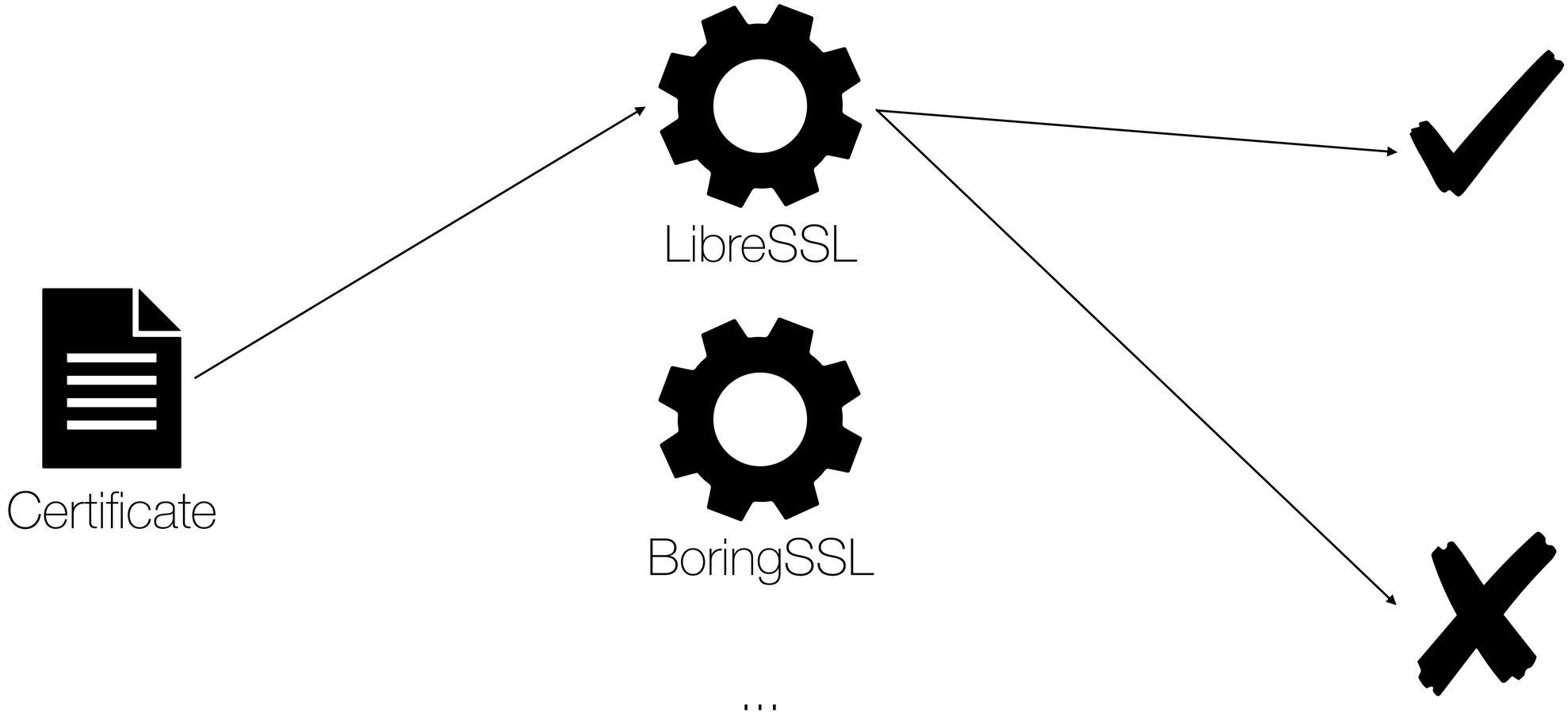
Certificate



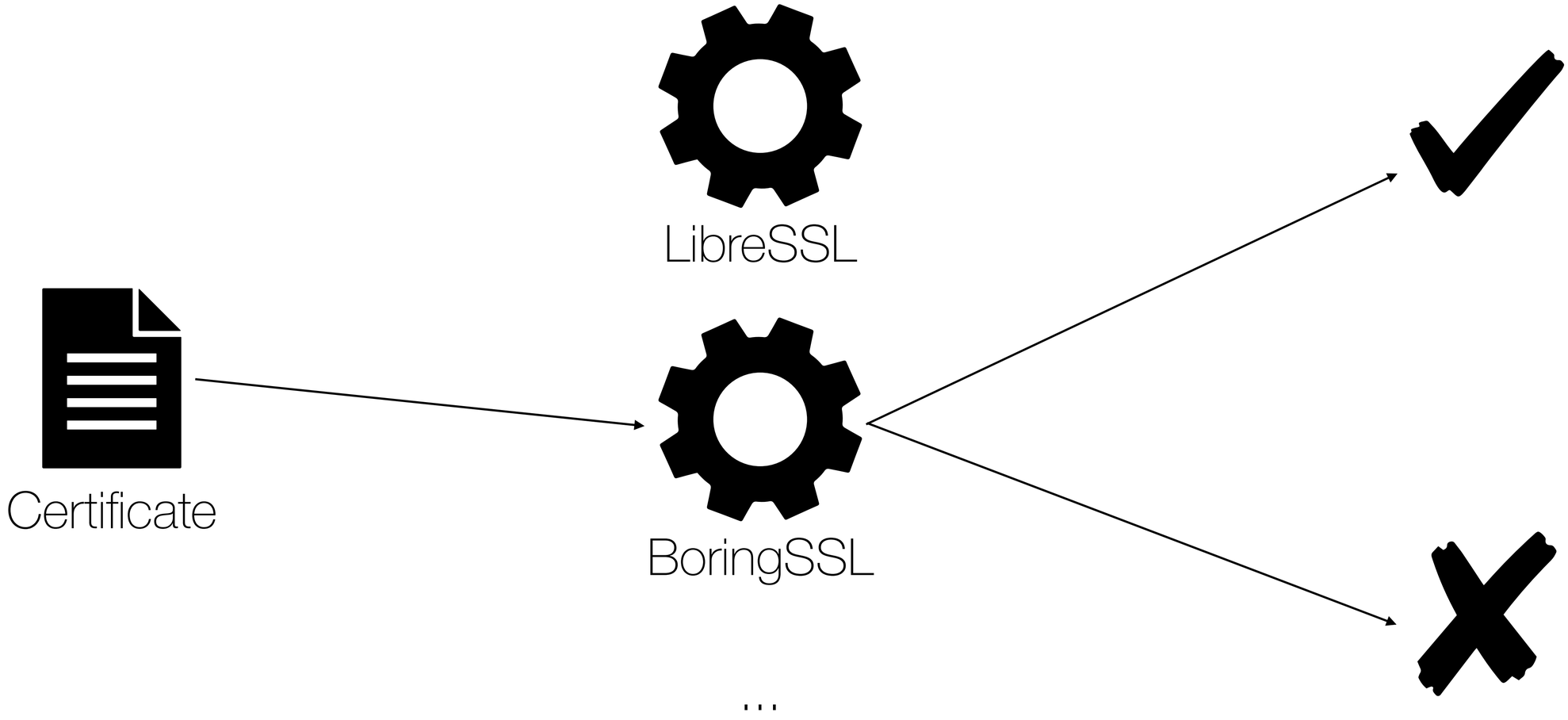
BoringSSL

...

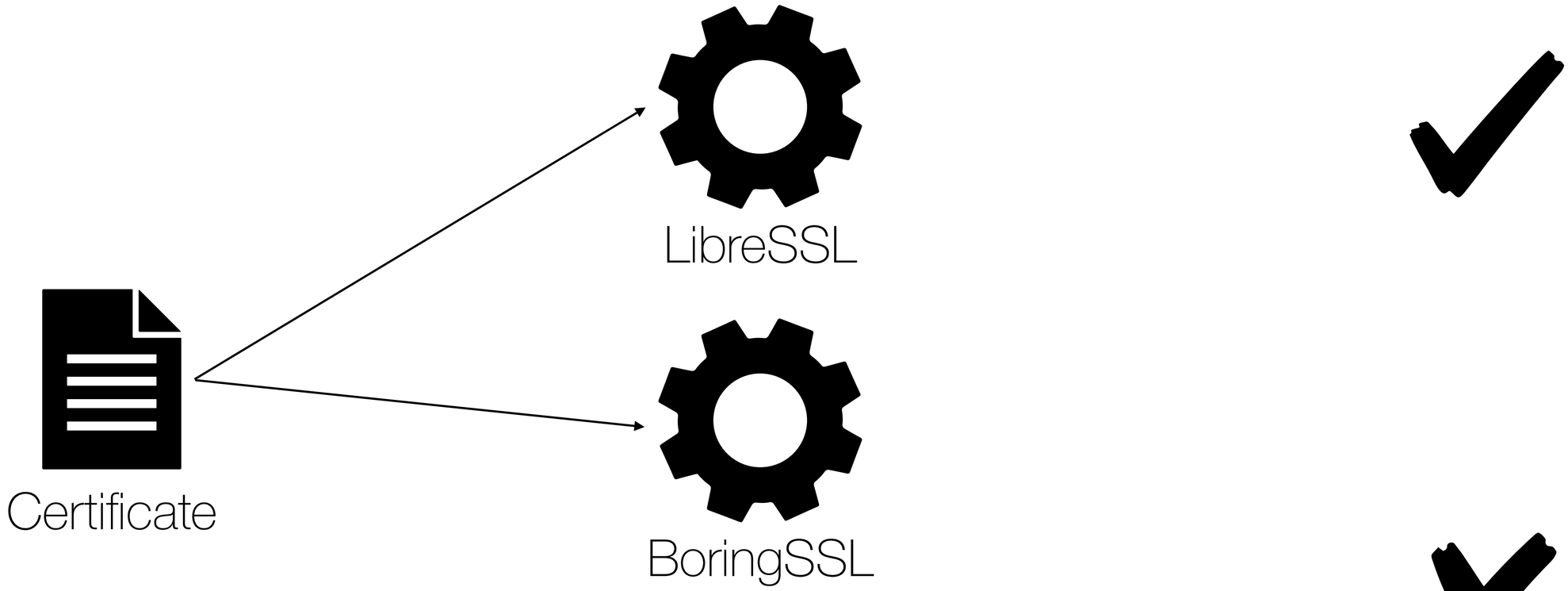
Various SSL/TLS implementations exist.



Various SSL/TLS implementations exist.

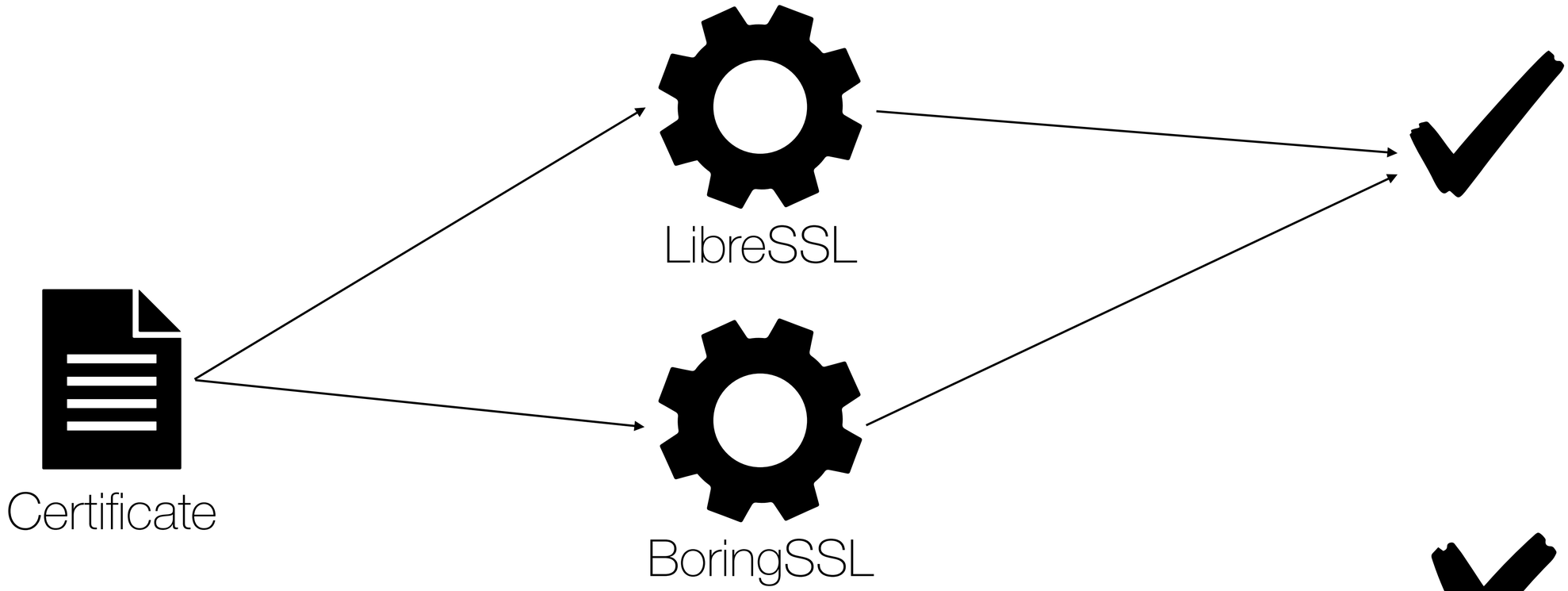


Various SSL/TLS implementations exist.



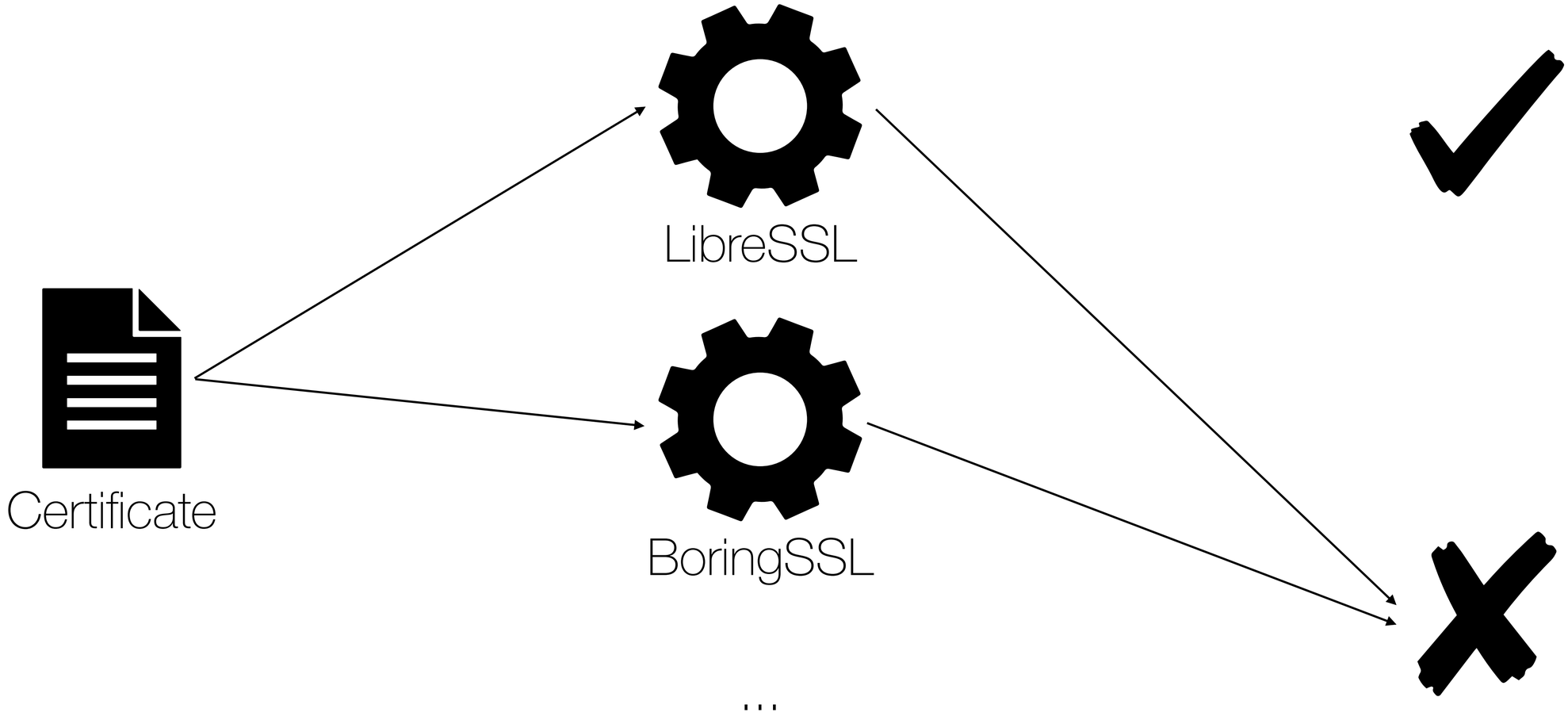
...

Various SSL/TLS implementations exist.

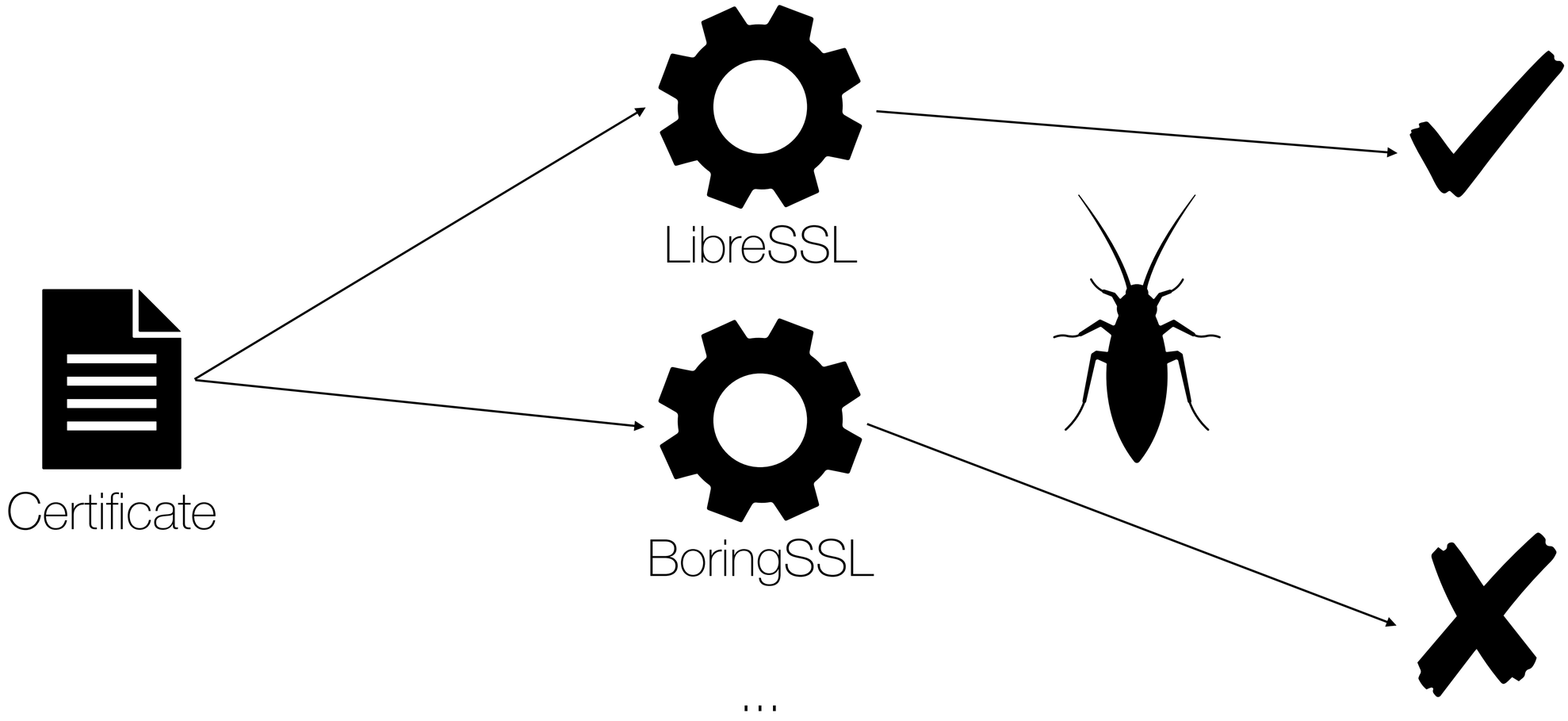


...

Various SSL/TLS implementations exist.

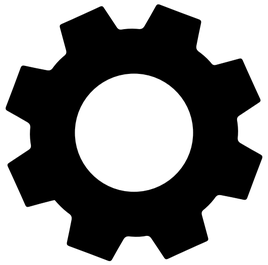


Various SSL/TLS implementations exist.

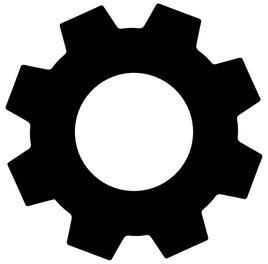


Various C compilers exist.

C Source Code



GCC



Clang

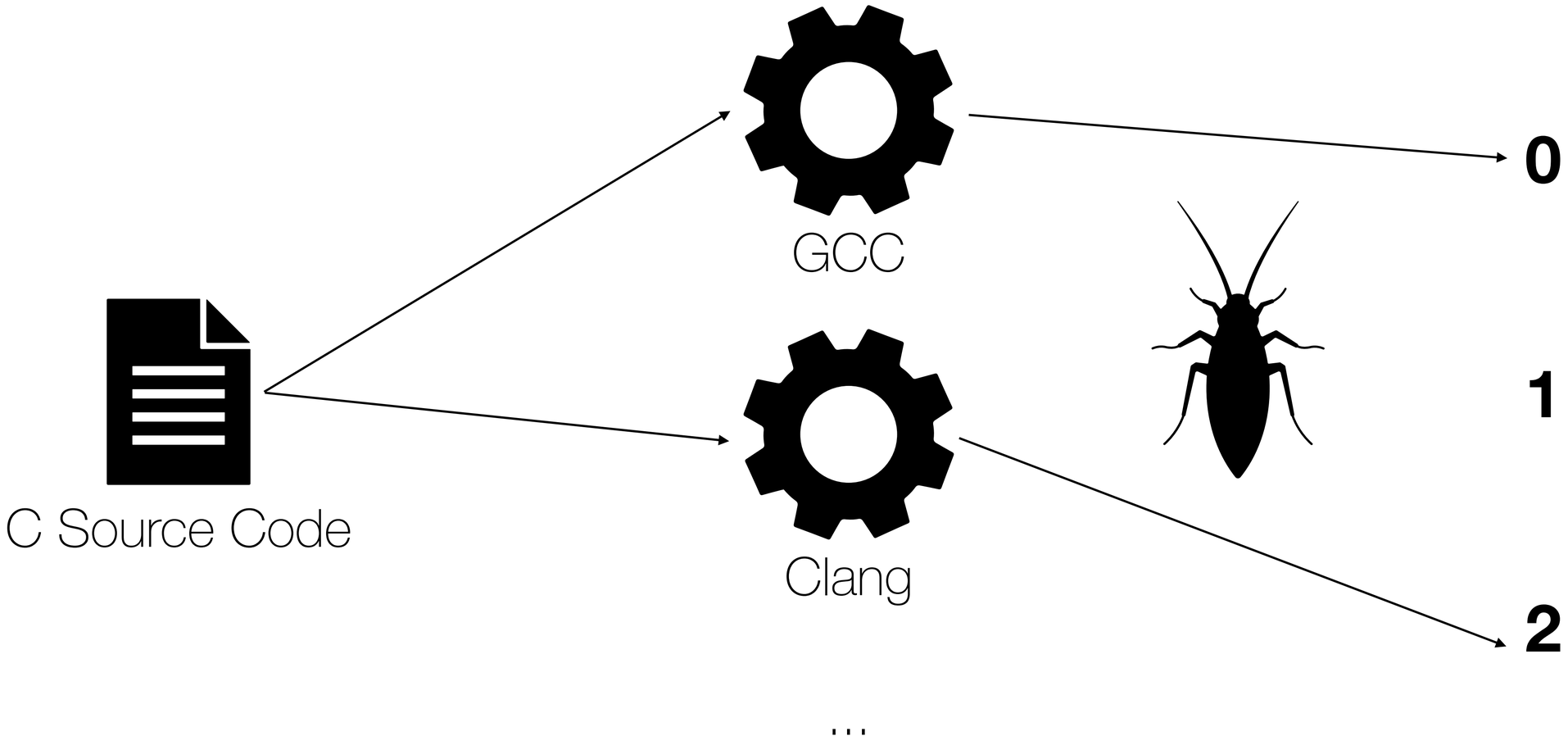
...

0

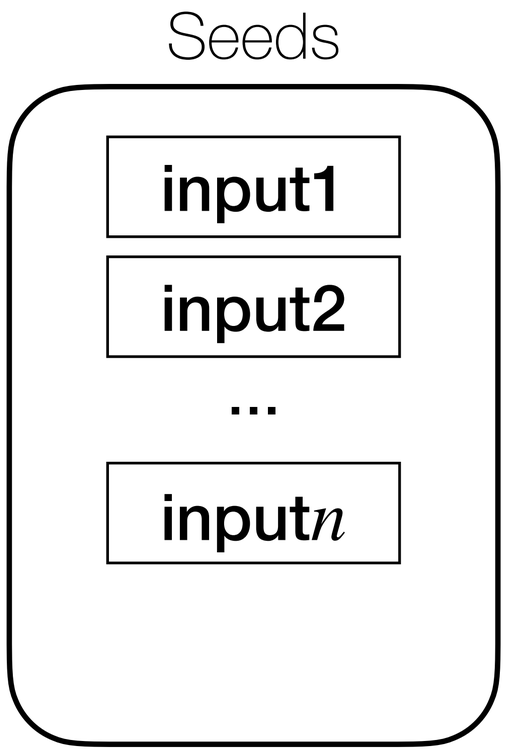
1

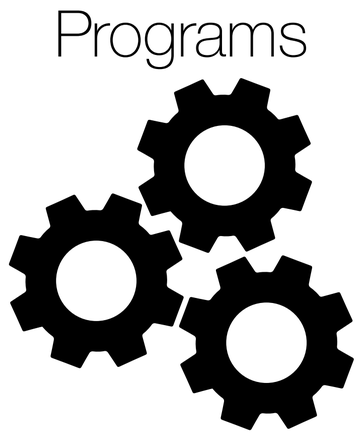
2

Various C compilers exist.

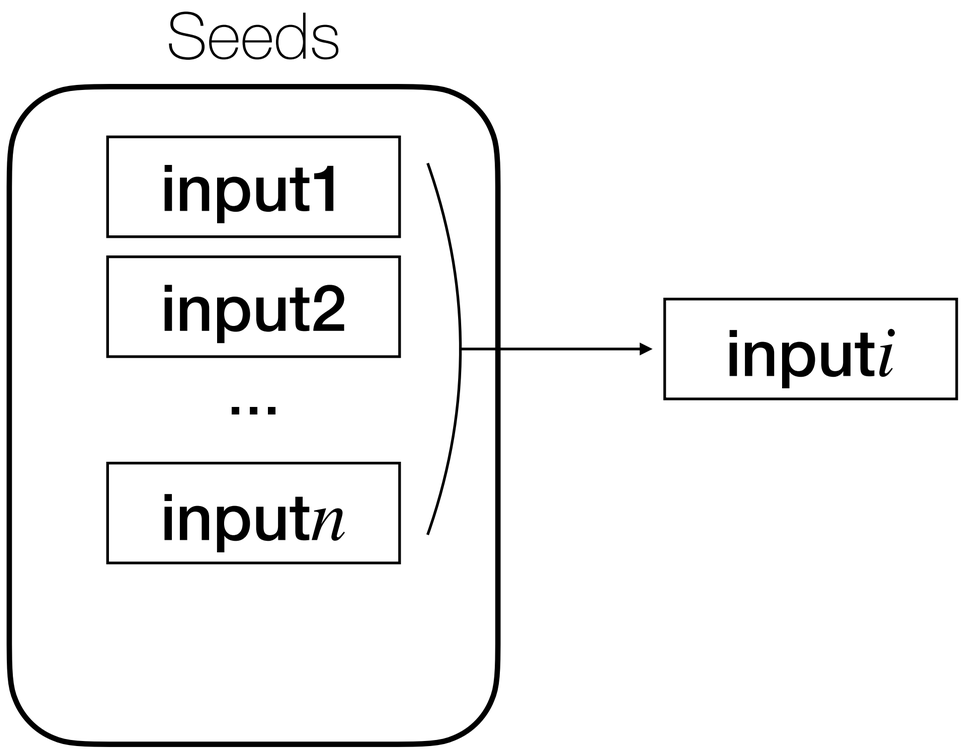


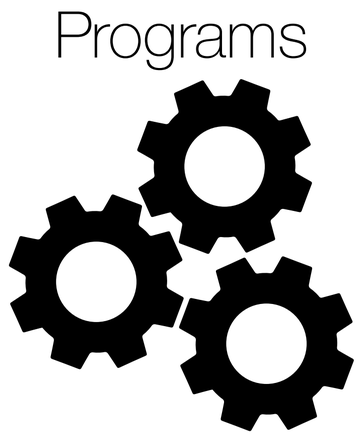
Differential testing runs multiple programs.



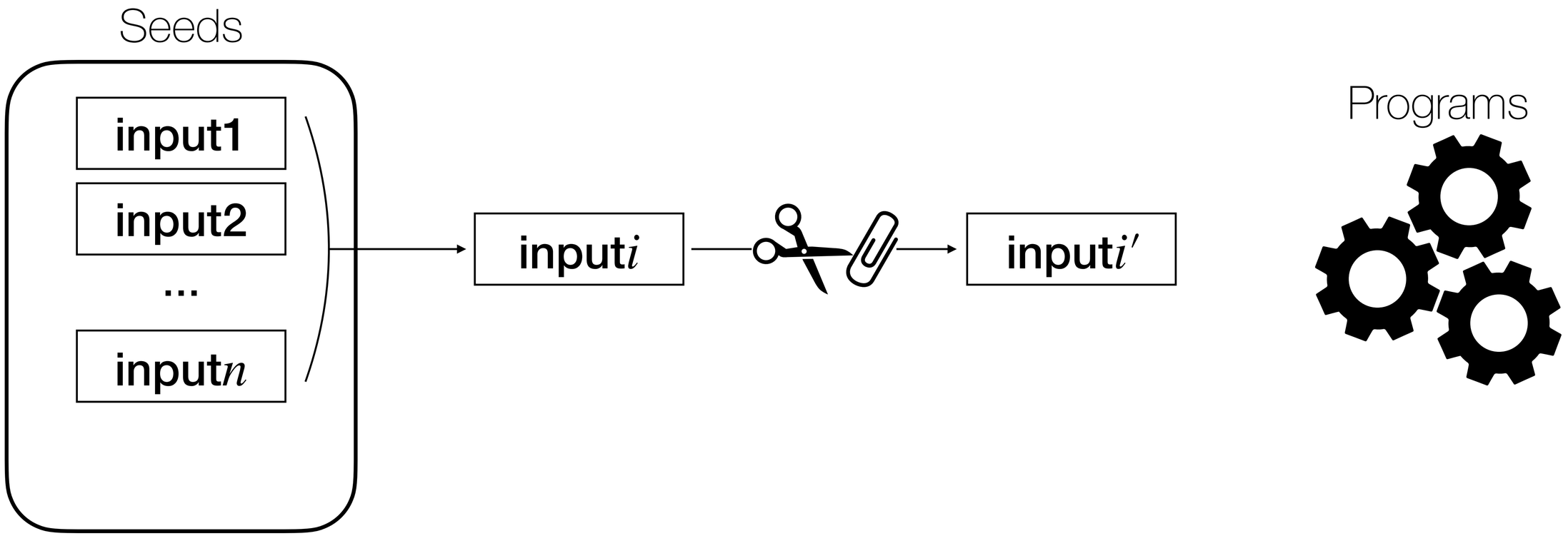


Differential testing runs multiple programs.

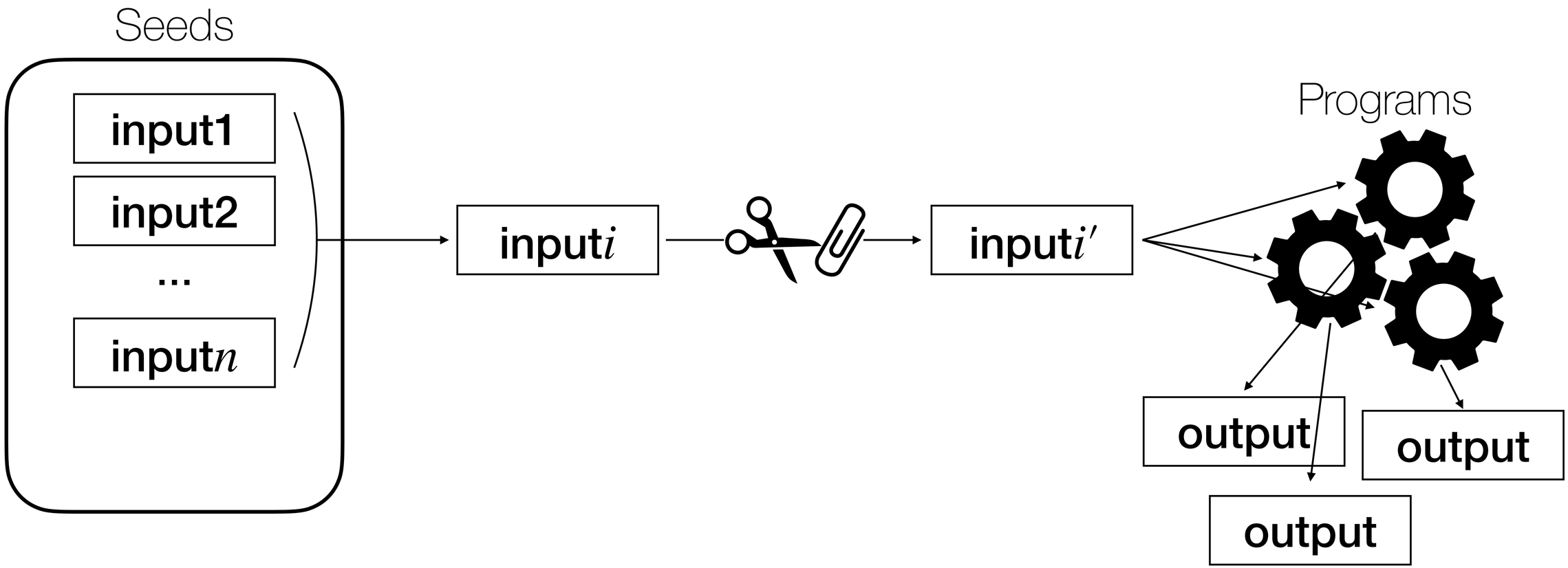




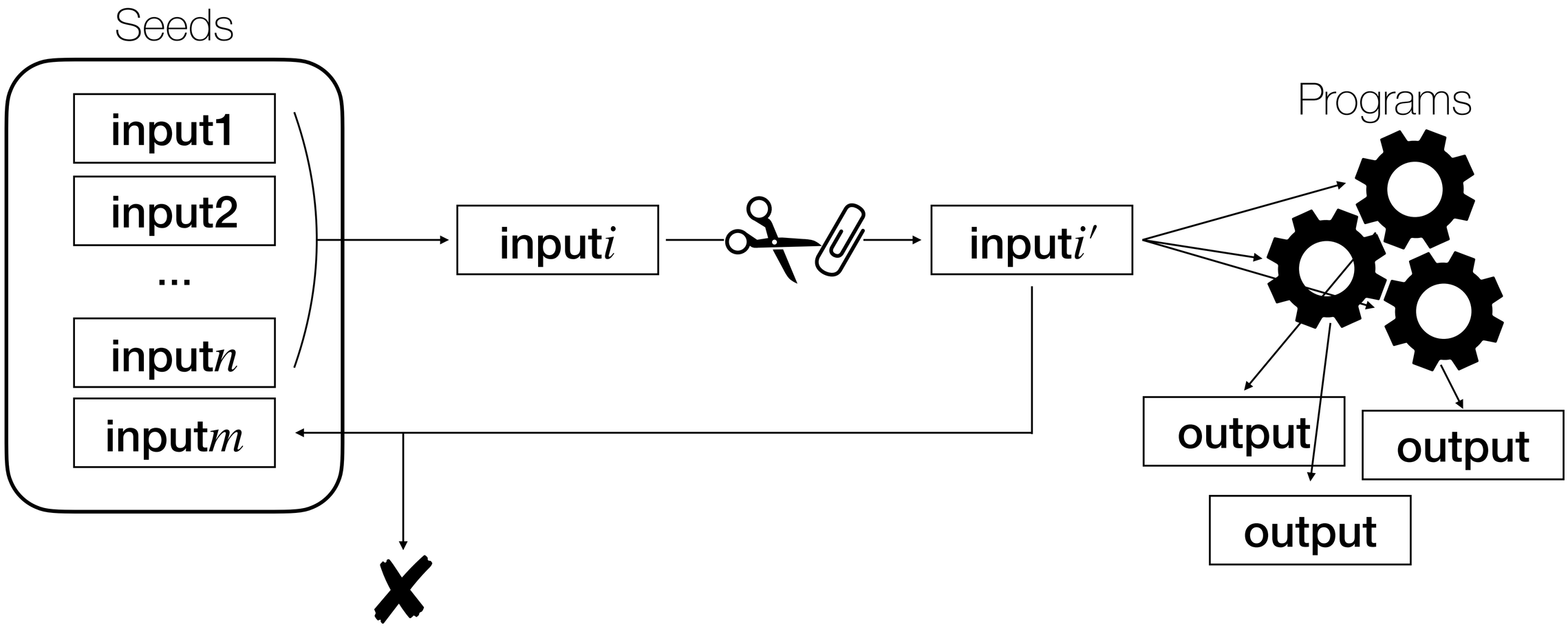
Differential testing runs multiple programs.



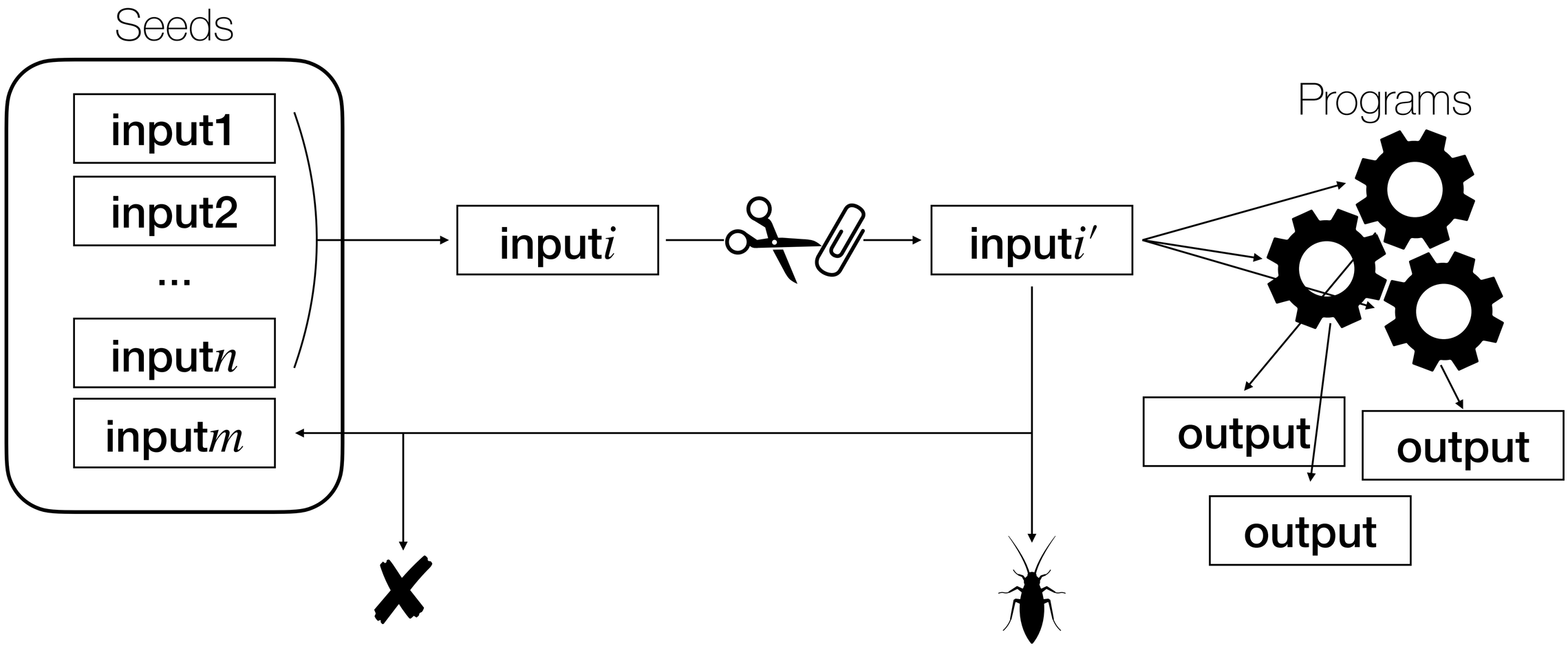
Differential testing runs multiple programs.



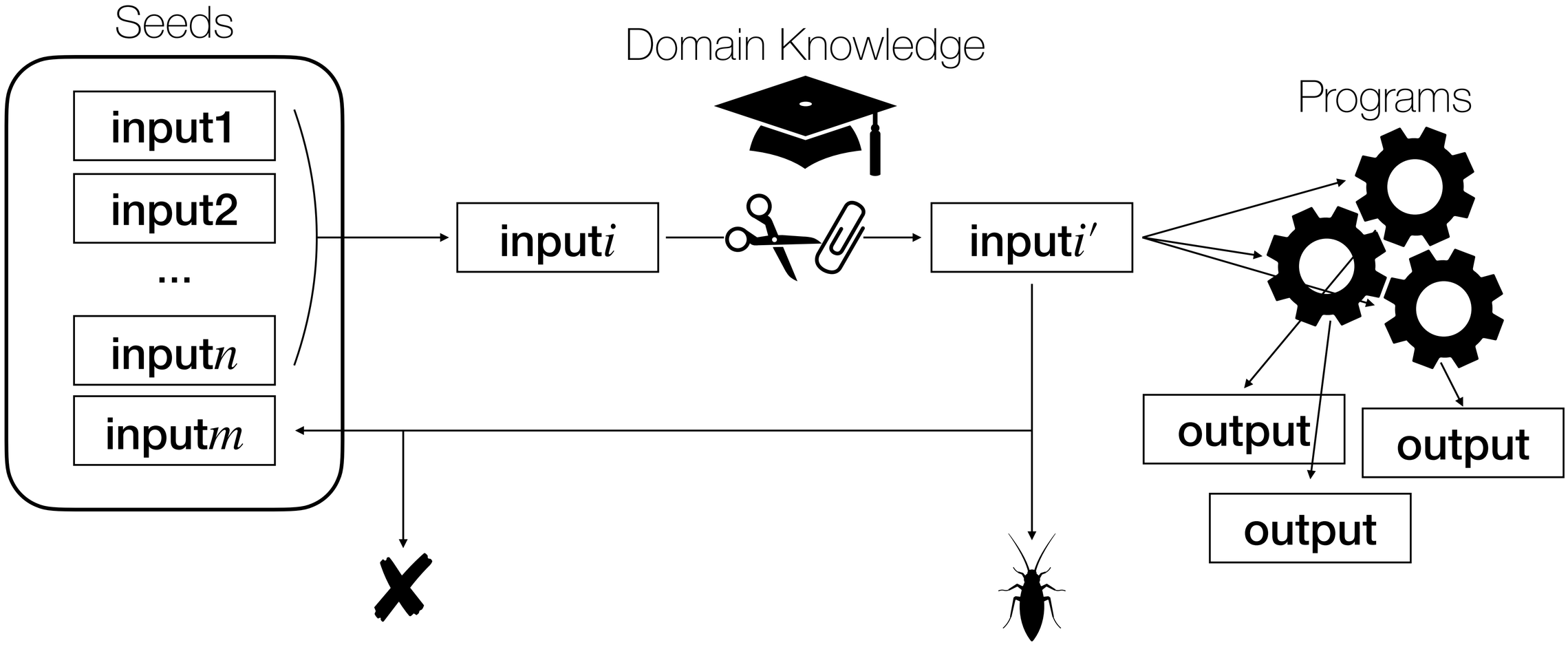
Differential testing runs multiple programs.



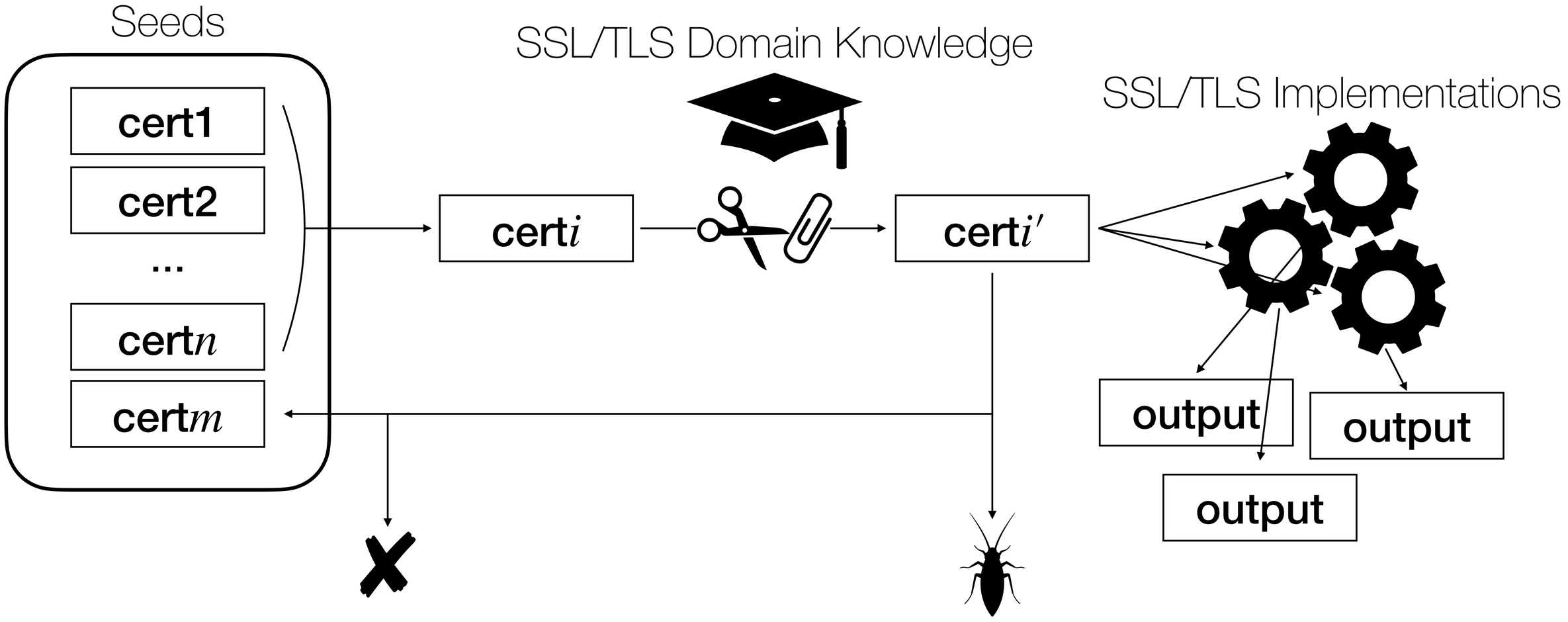
Differential testing runs multiple programs.



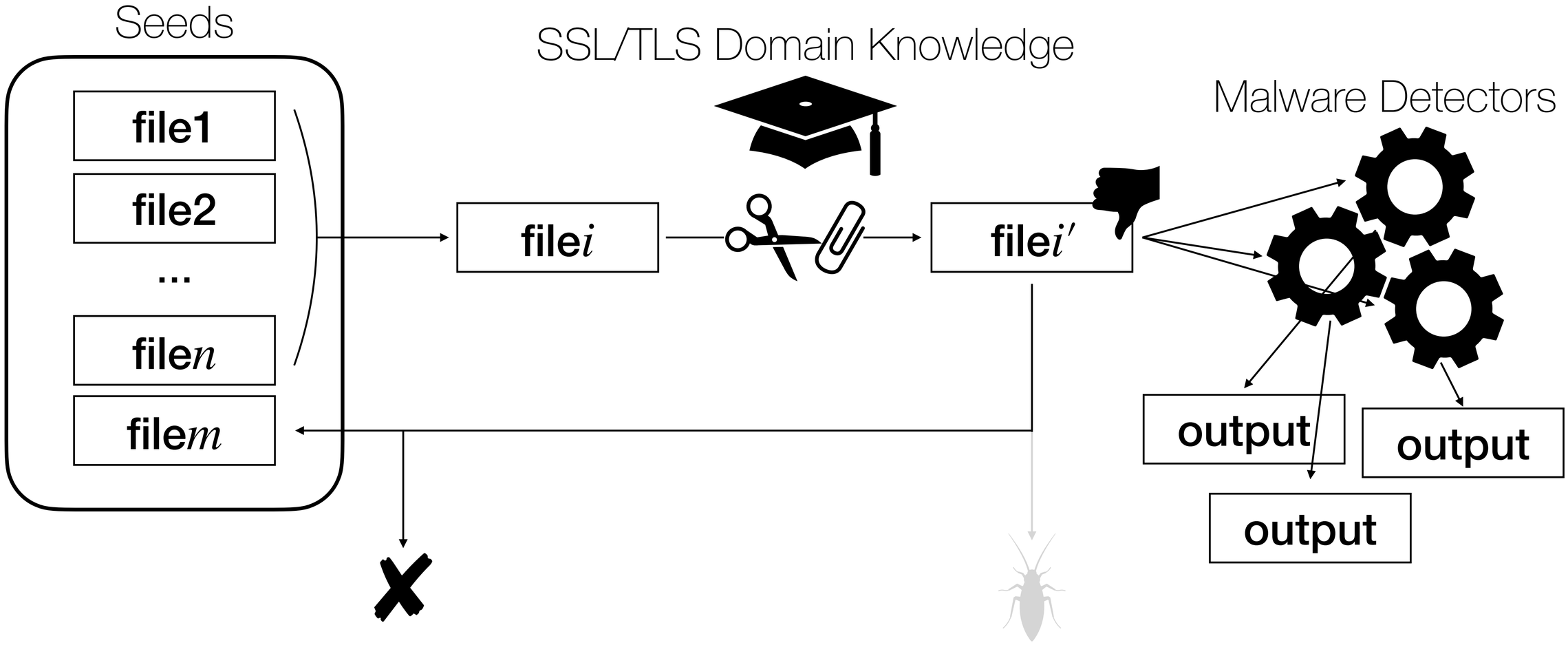
Existing differential testing frameworks are domain-specific.



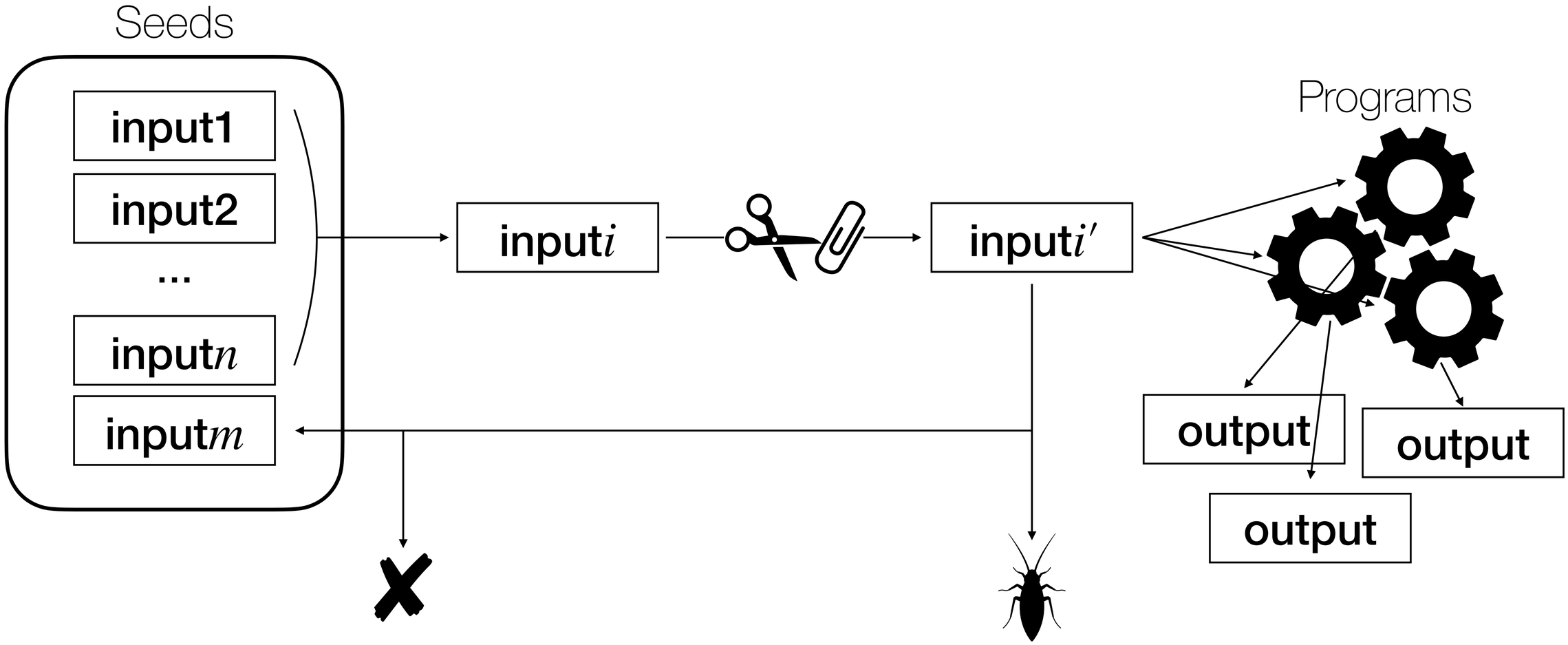
Existing differential testing frameworks are domain-specific.



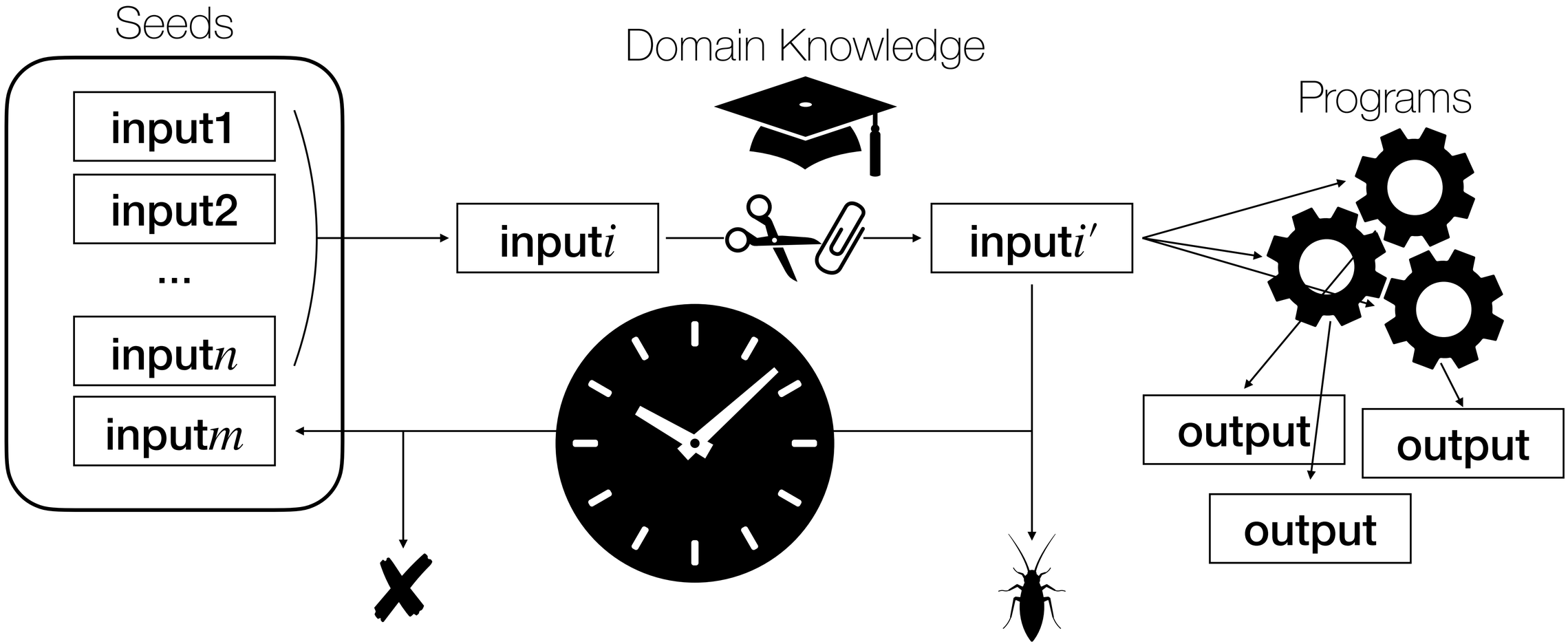
Existing differential testing frameworks are domain-specific.



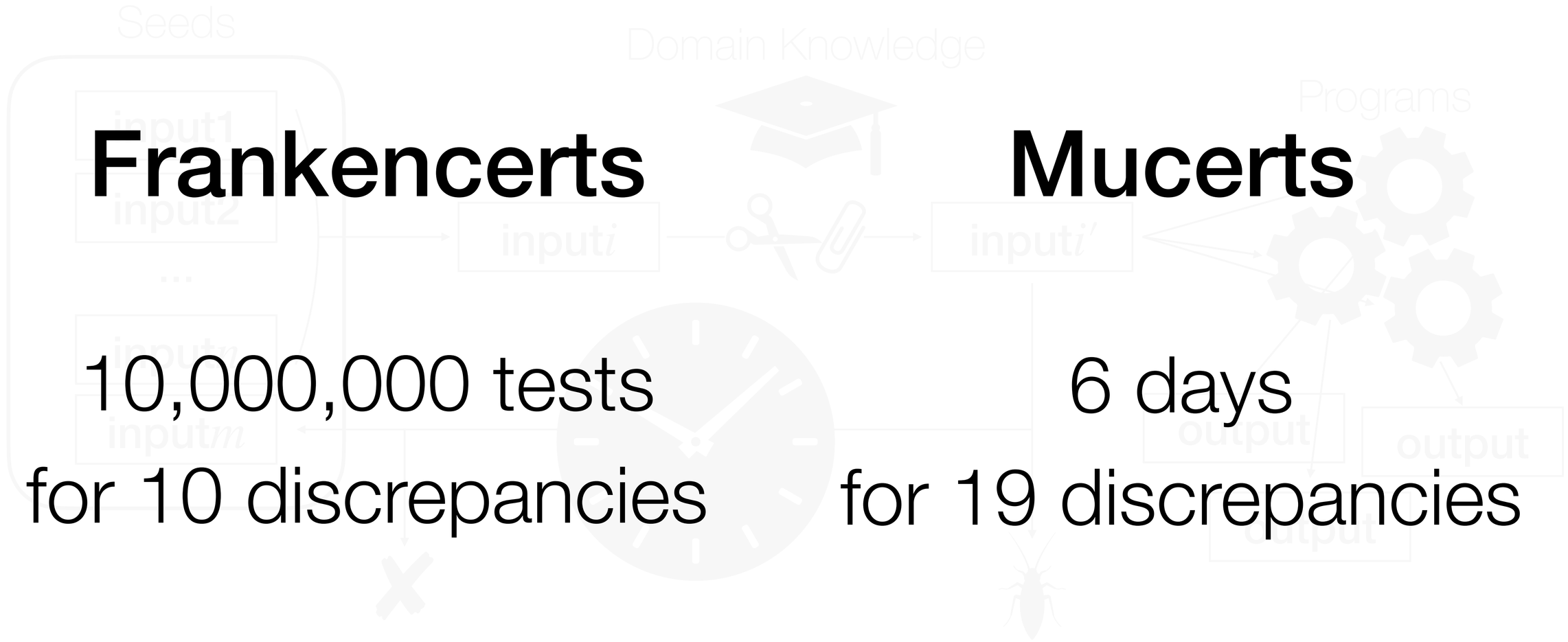
A domain-independent differential testing framework is required.



Existing domain-specific differential testing frameworks are inefficient.



Existing domain-specific differential testing frameworks are inefficient.



Frankencerts

10,000,000 tests
for 10 discrepancies

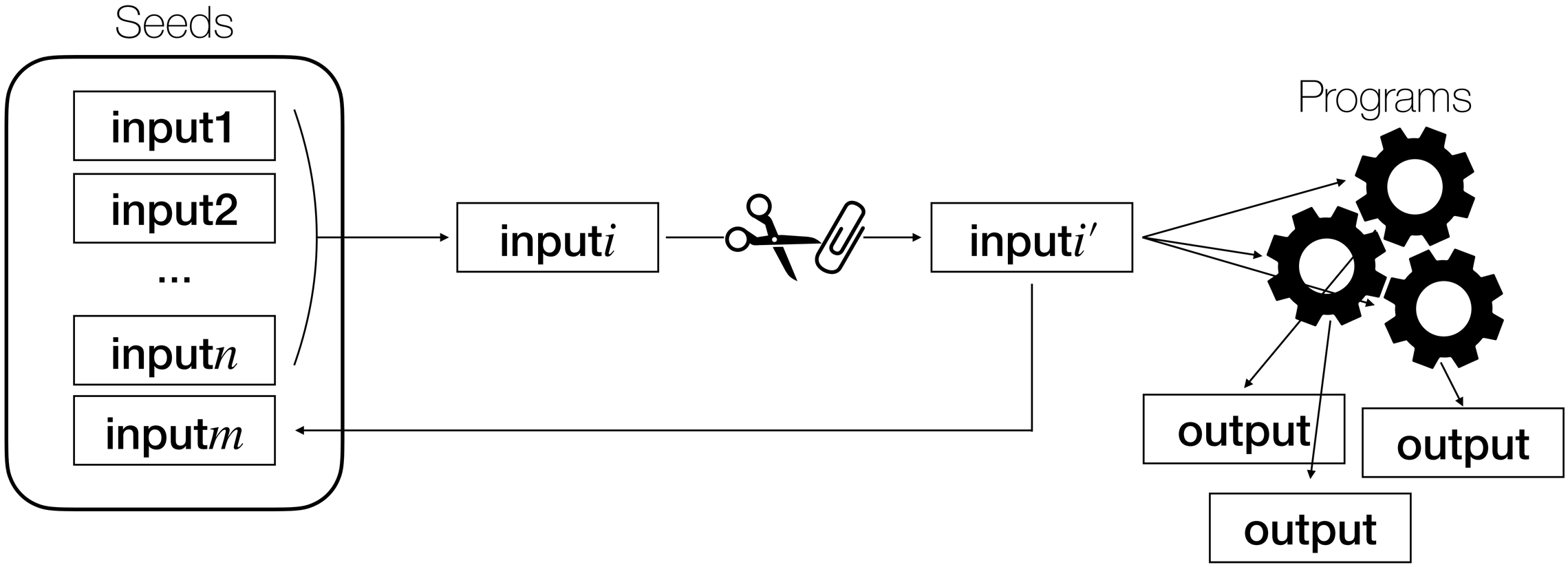
Mucerts

6 days
for 19 discrepancies

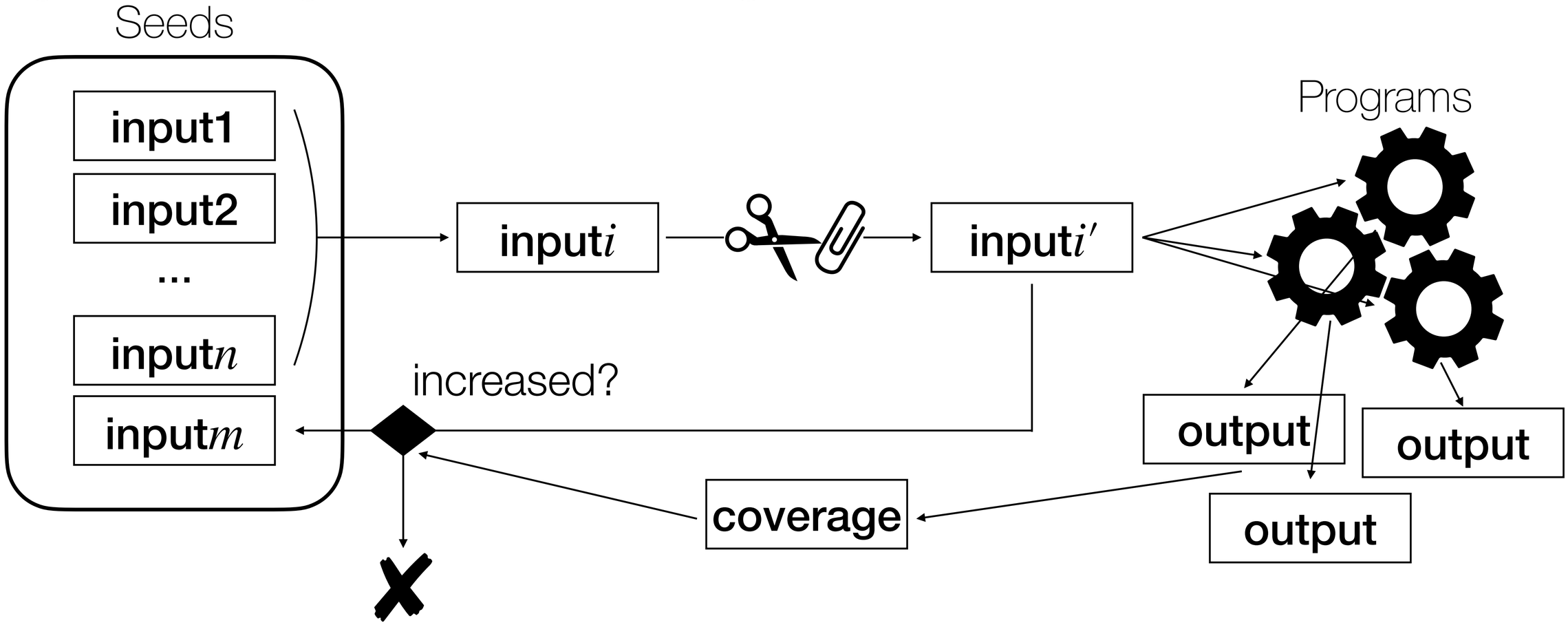
Nezha

is a domain-independent differential testing framework outperforming existing frameworks.

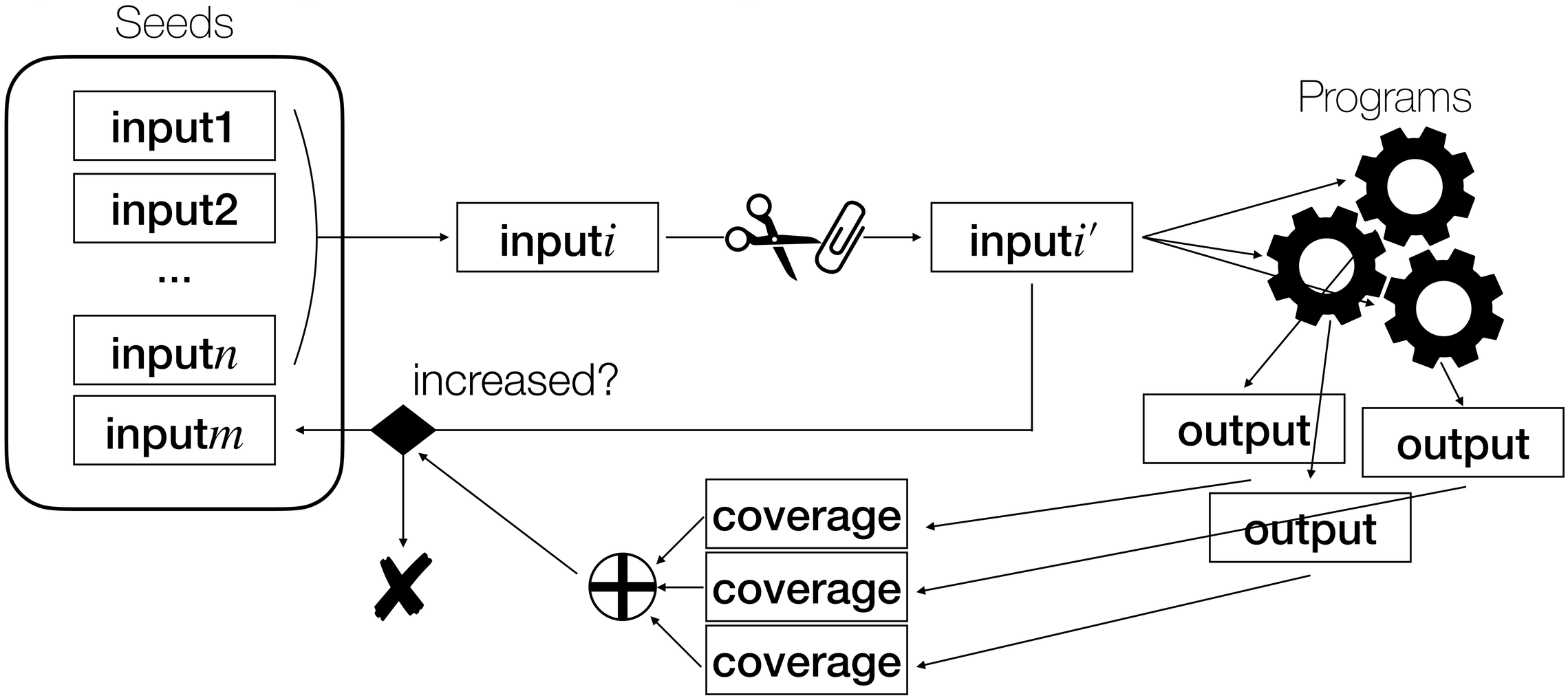
Existing differential testing frameworks ignore asymmetry across programs.



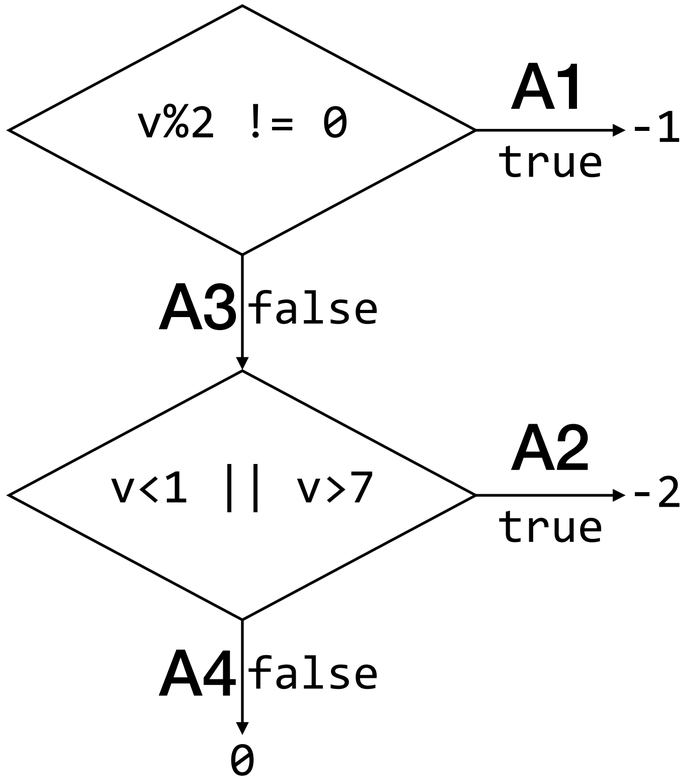
Existing differential testing frameworks ignore asymmetry across programs.

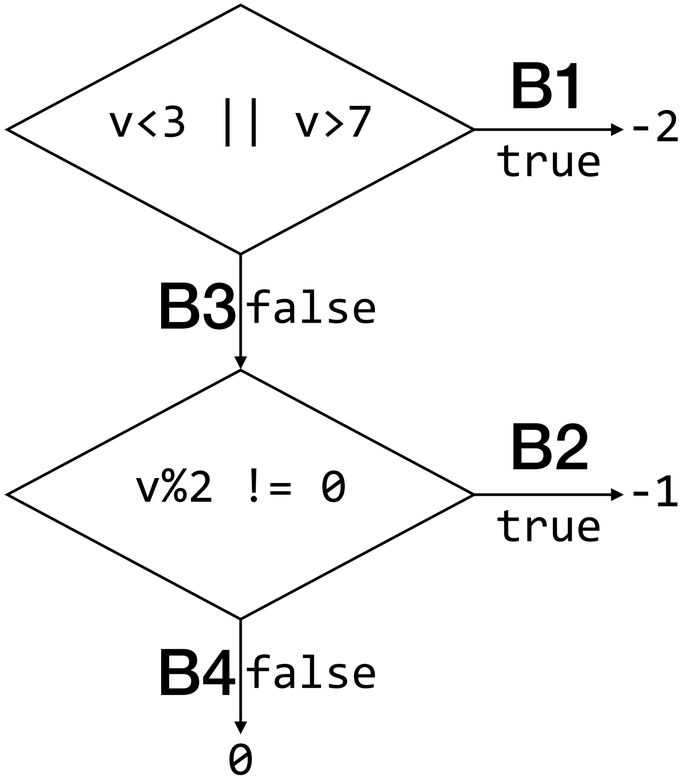


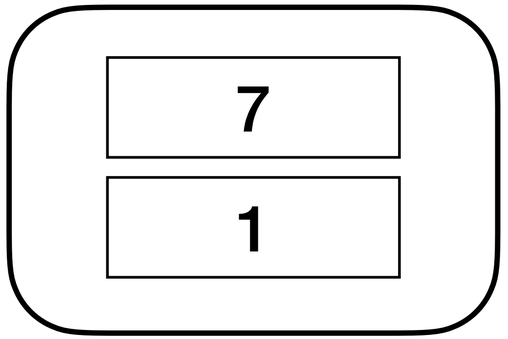
Existing differential testing frameworks ignore asymmetry across programs.

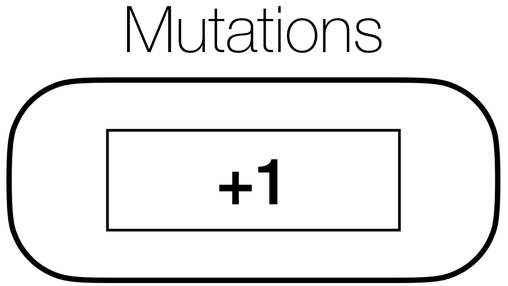


The strategy maximizing the global coverage often misses interesting inputs.

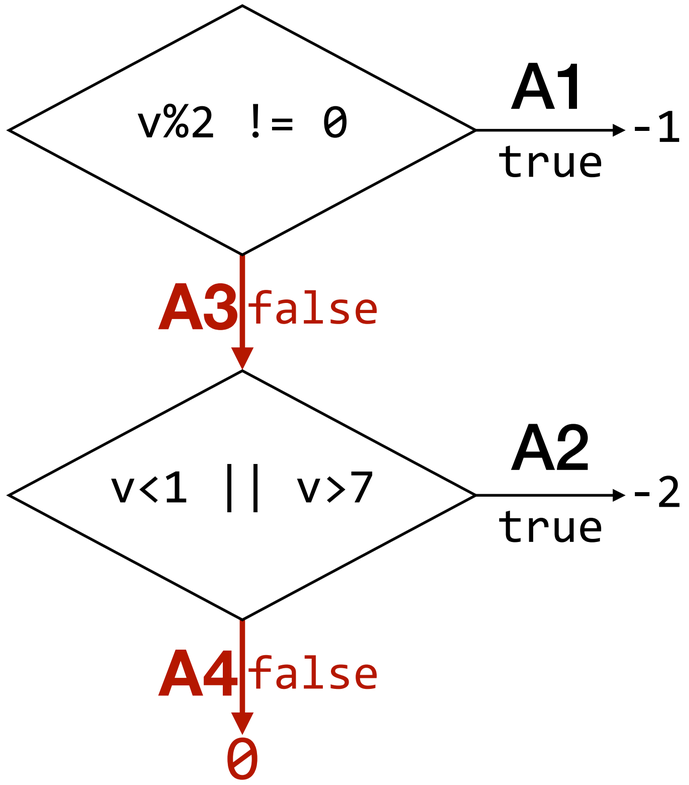


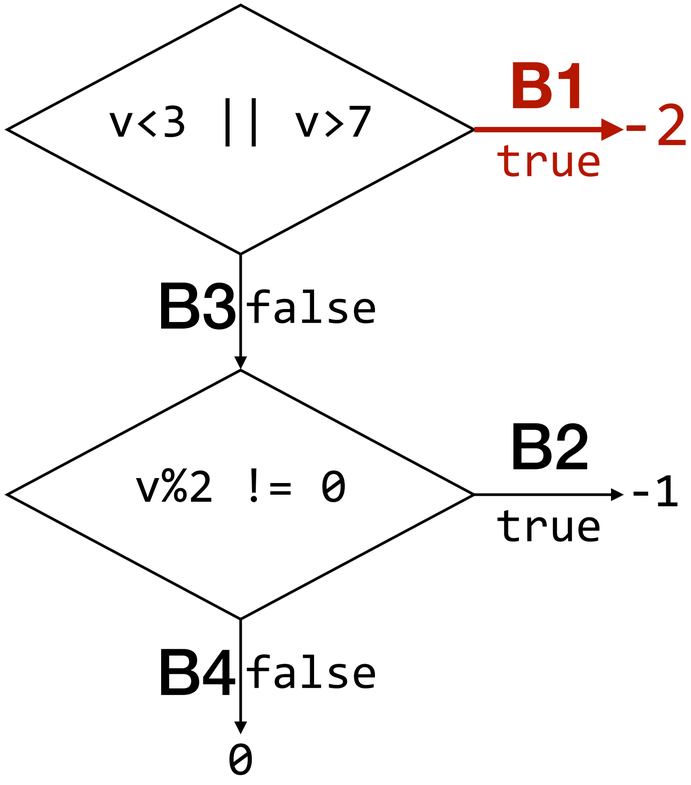


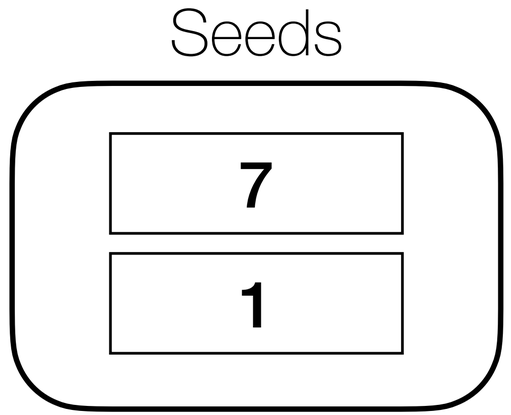


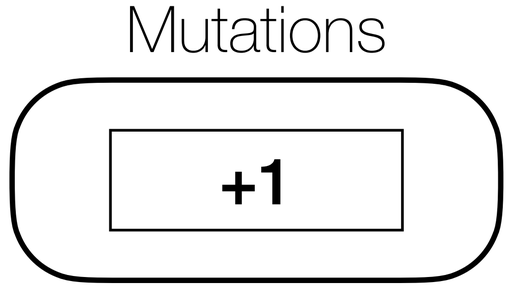


The strategy maximizing the global coverage often misses interesting inputs.



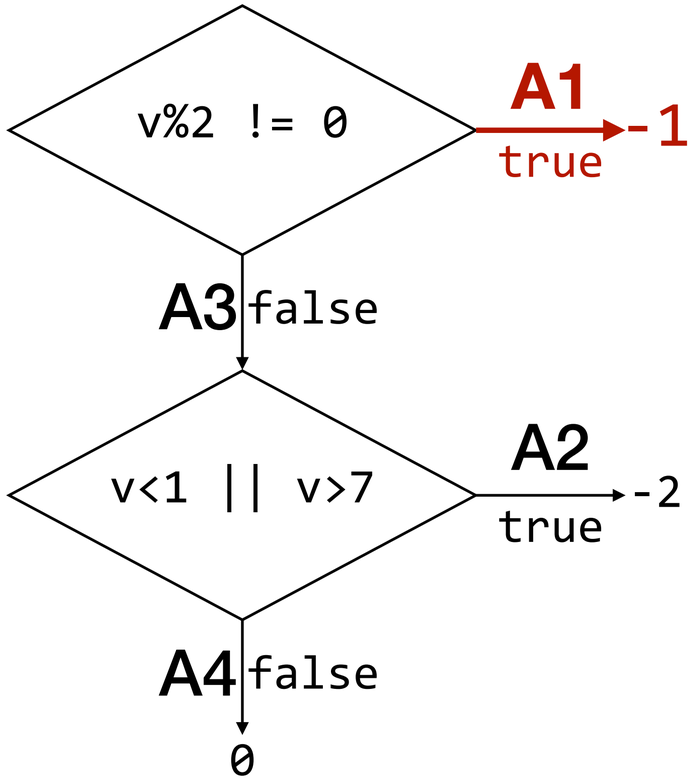


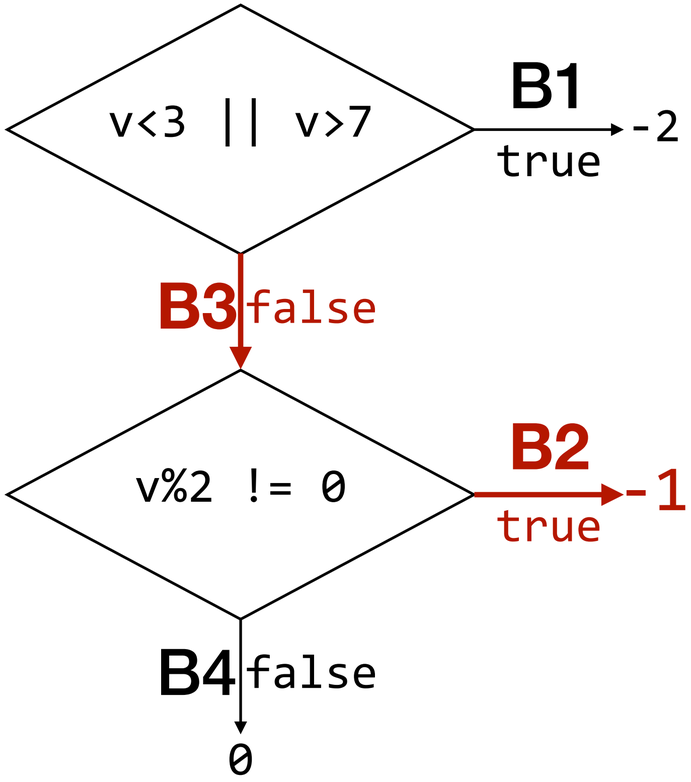


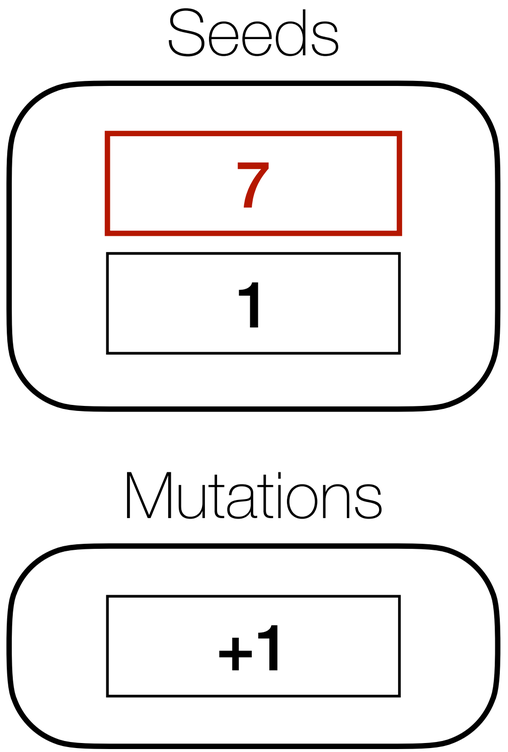


Differs on input 2

The strategy maximizing the global coverage often misses interesting inputs.

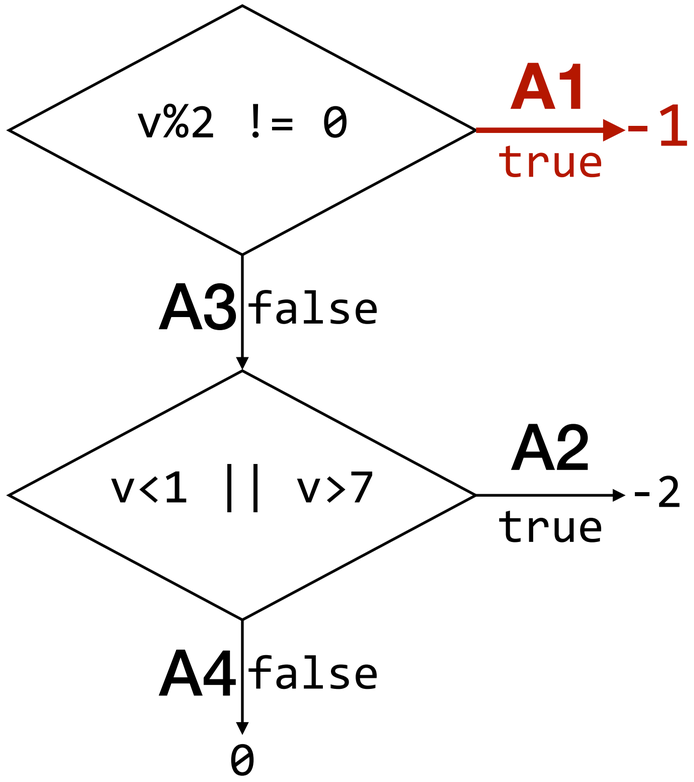


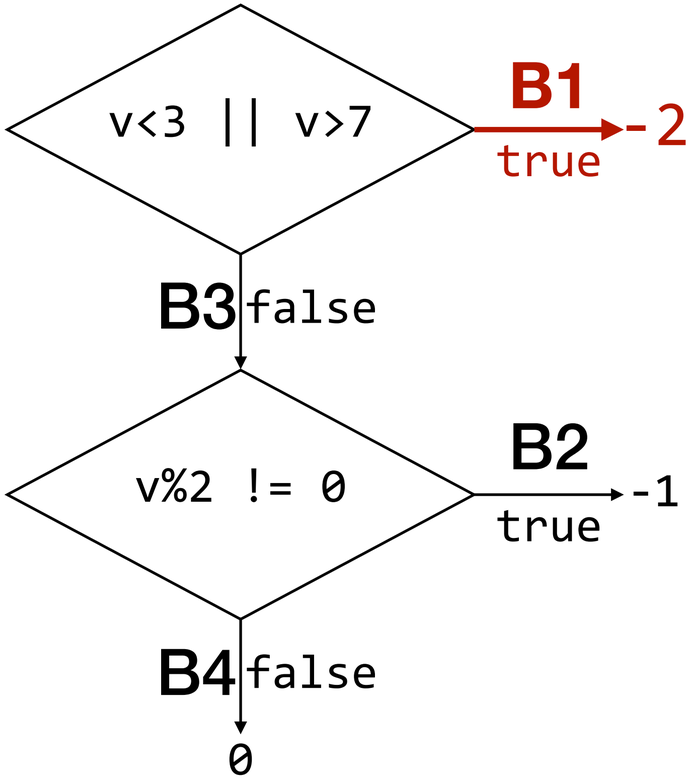


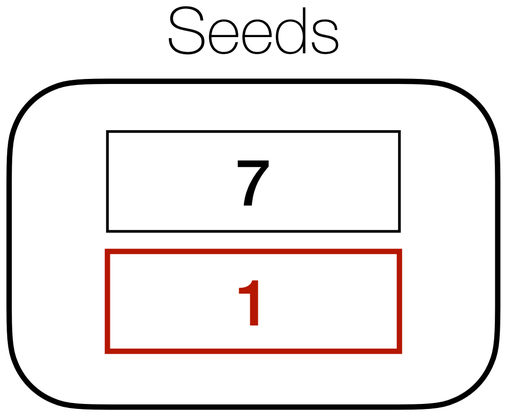


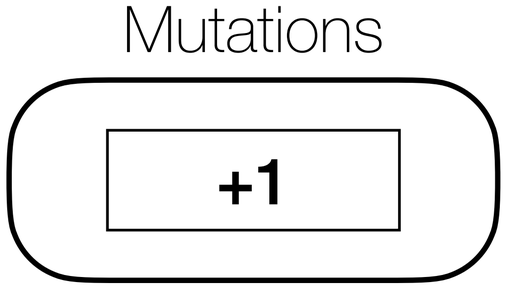
coverage = 3

The strategy maximizing the global coverage often misses interesting inputs.



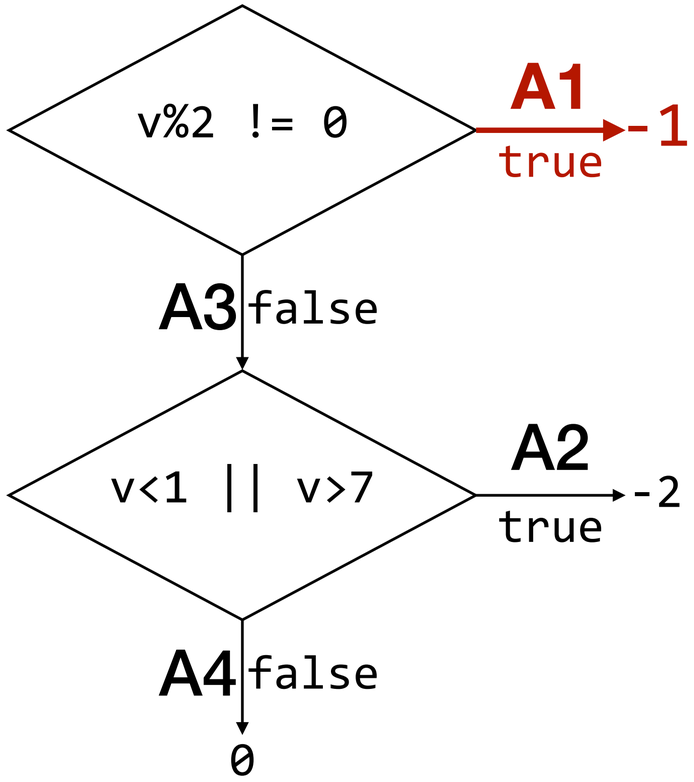


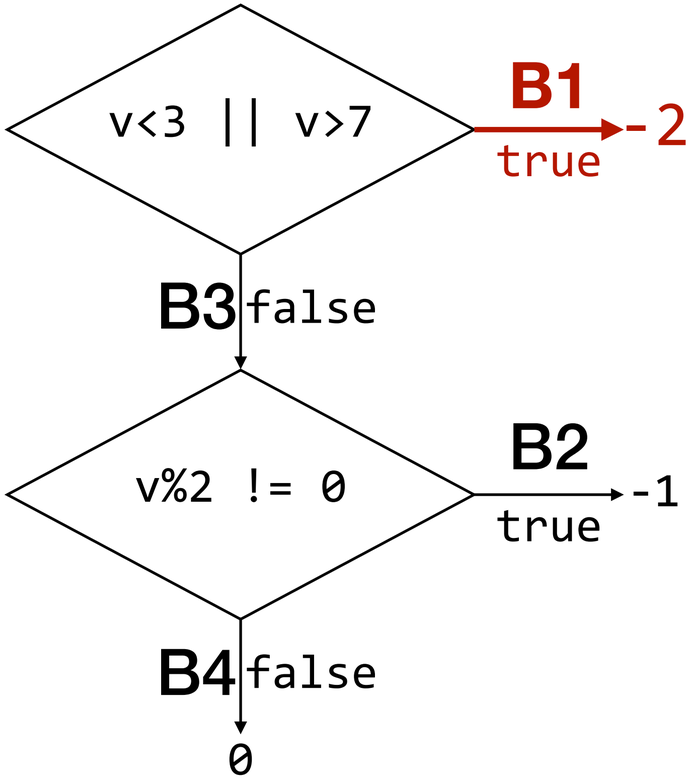


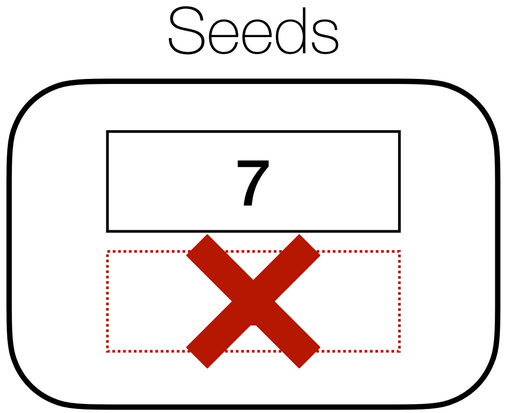


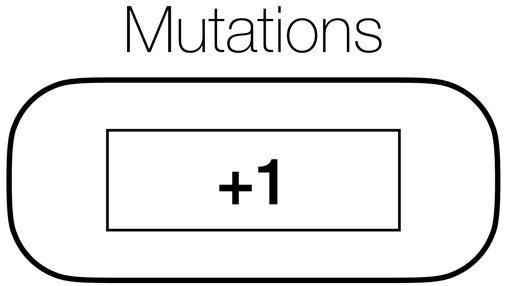
coverage = 2

The strategy maximizing the global coverage often misses interesting inputs.



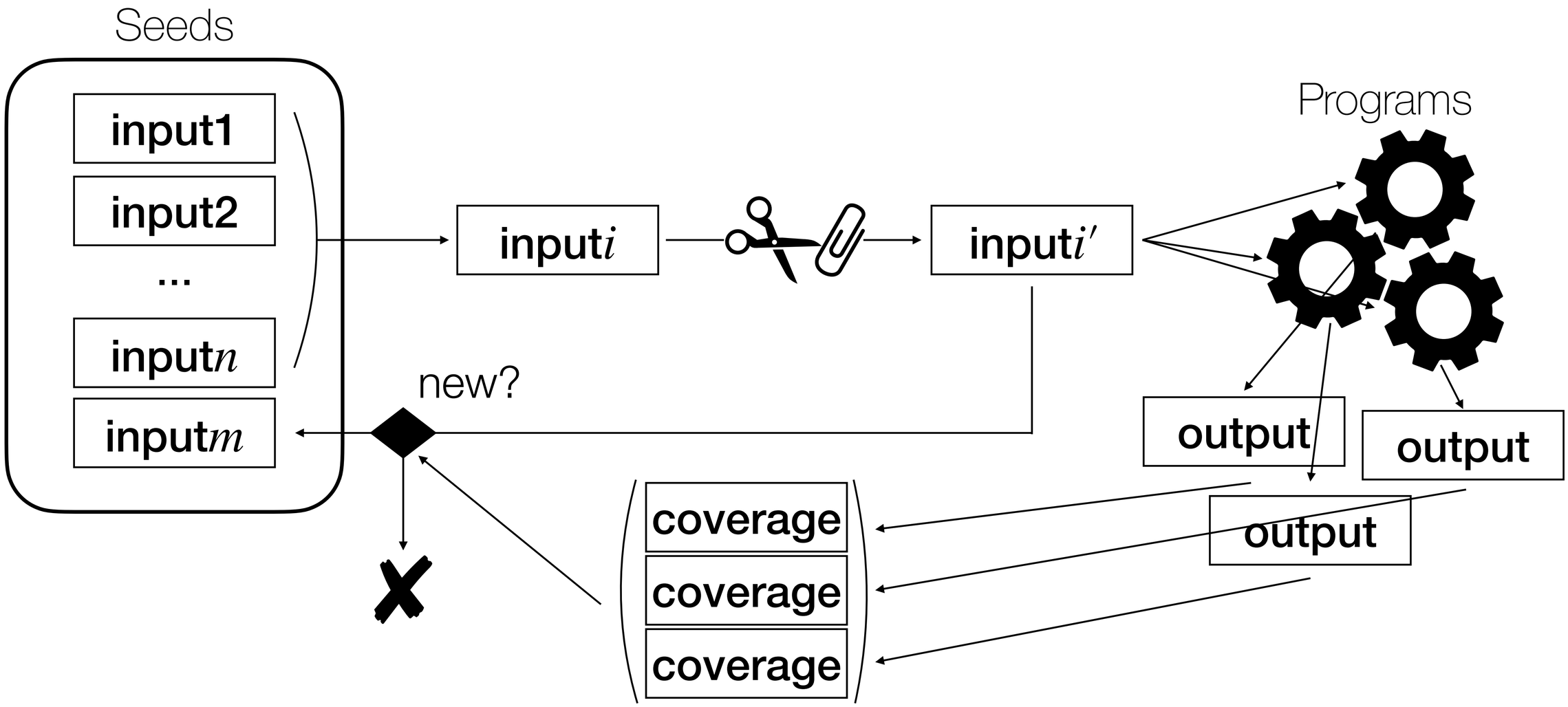






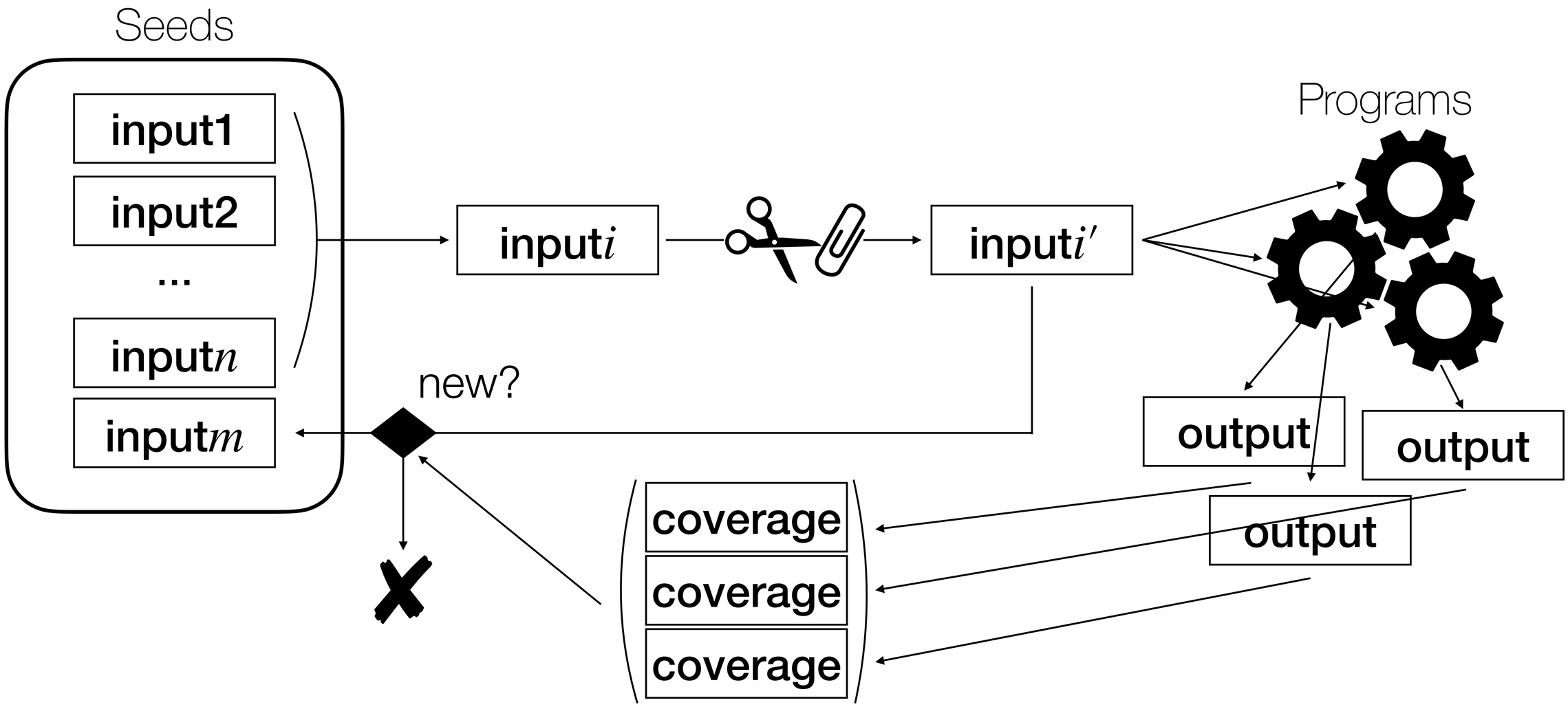
coverage = 2 < 3

NEZHA introduces δ-diversity.

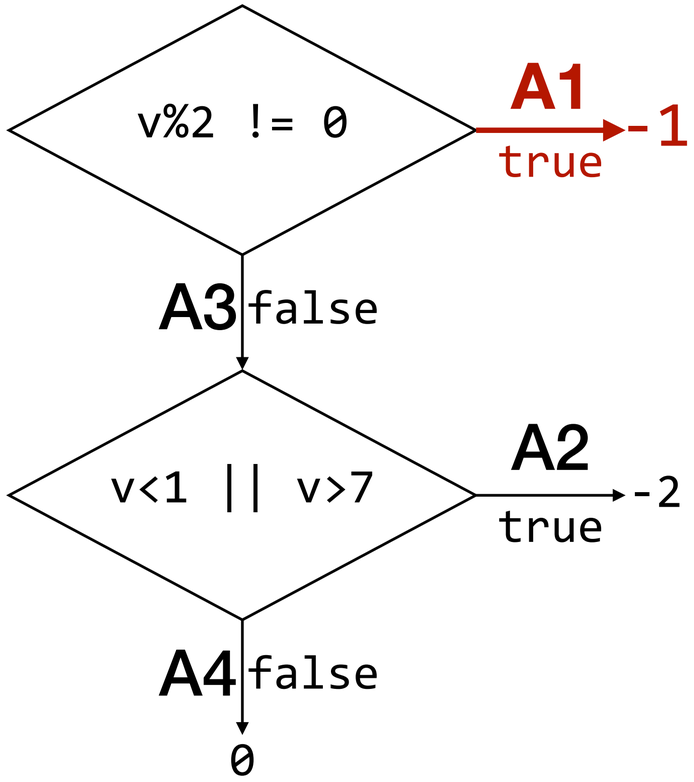


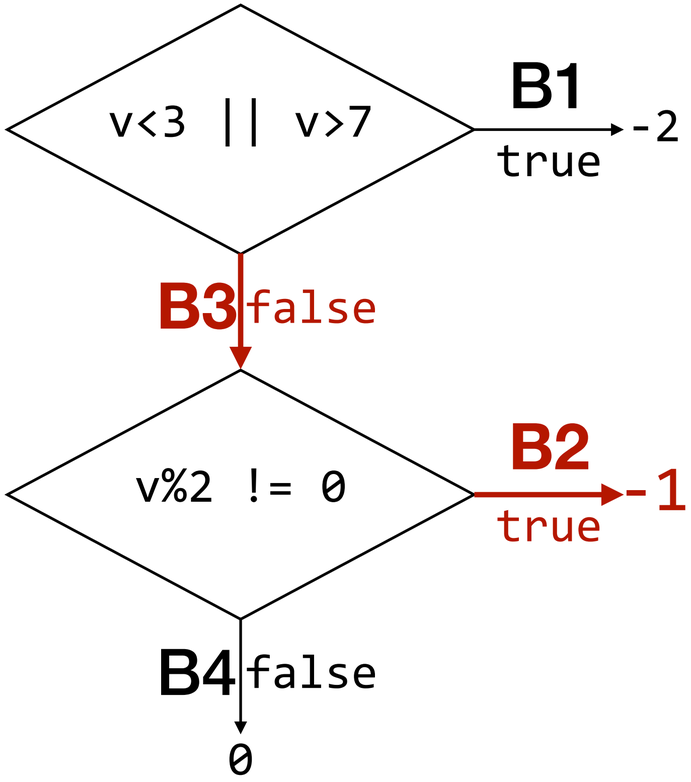
NEZHA introduces δ-diversity.

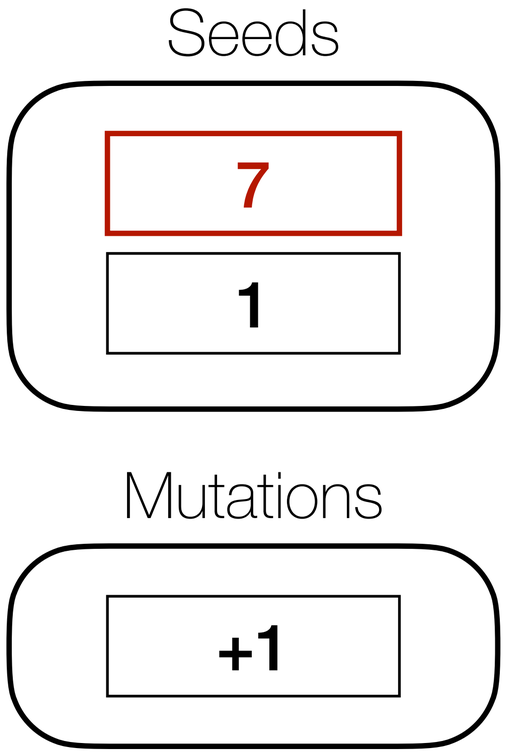
- Fine path δ-diversity
- Coarse path δ-diversity
- Output δ-diversity



Fine path δ-diversity considers tuples of sets.

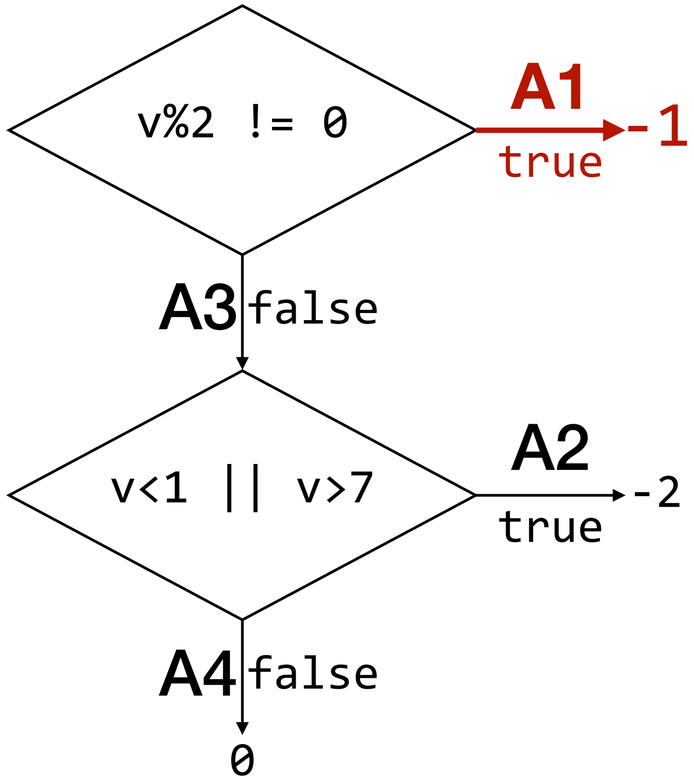


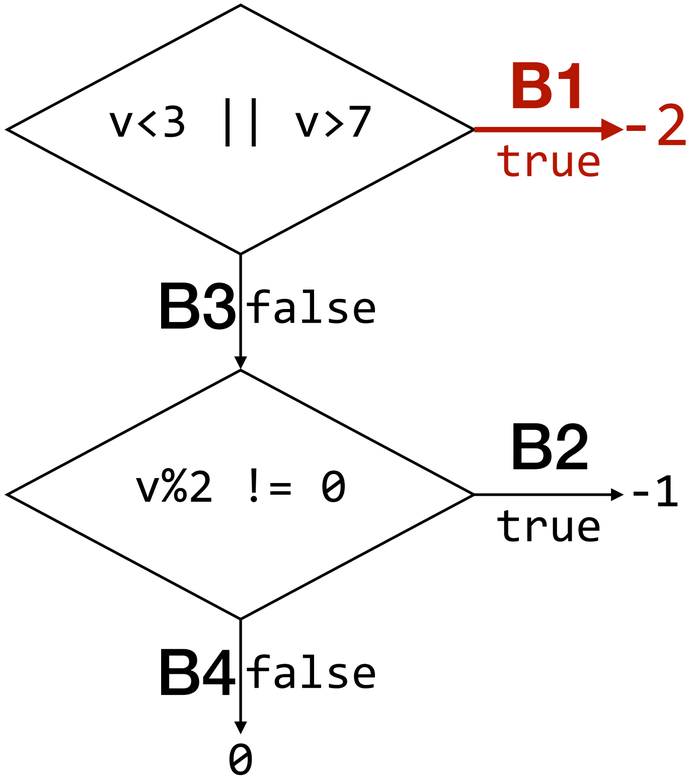


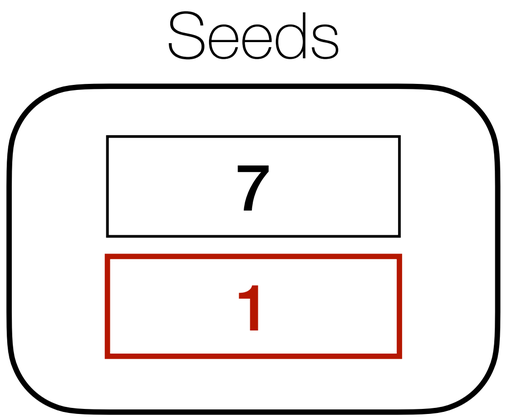


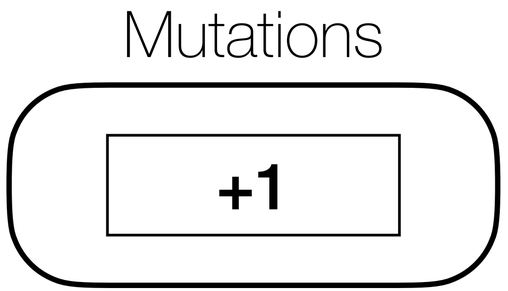
tuple = ({A1}, {B2, B3})

Fine path δ-diversity considers tuples of sets.



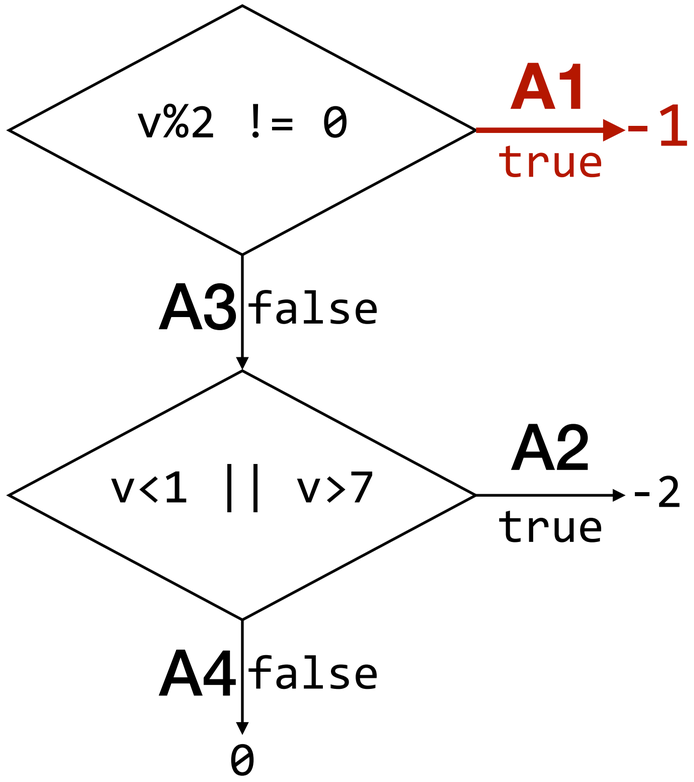


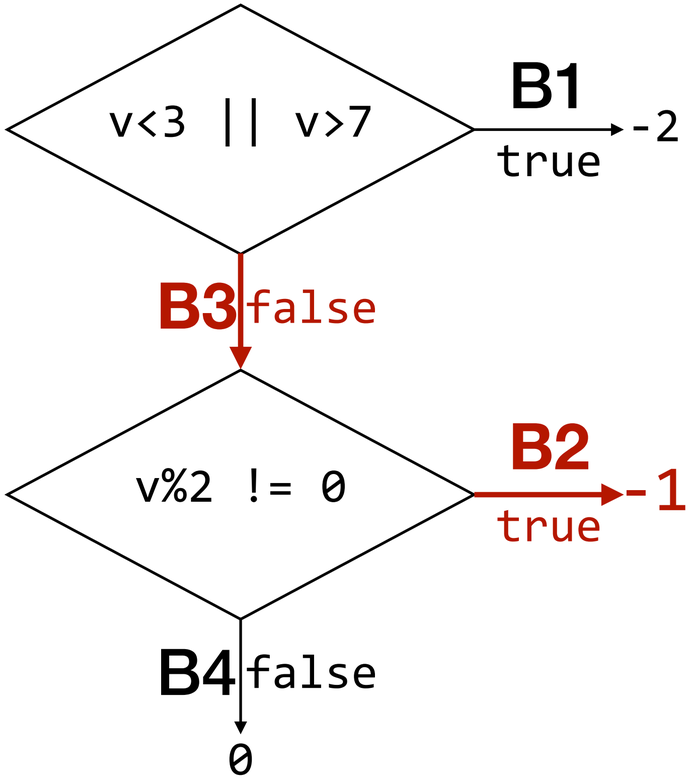


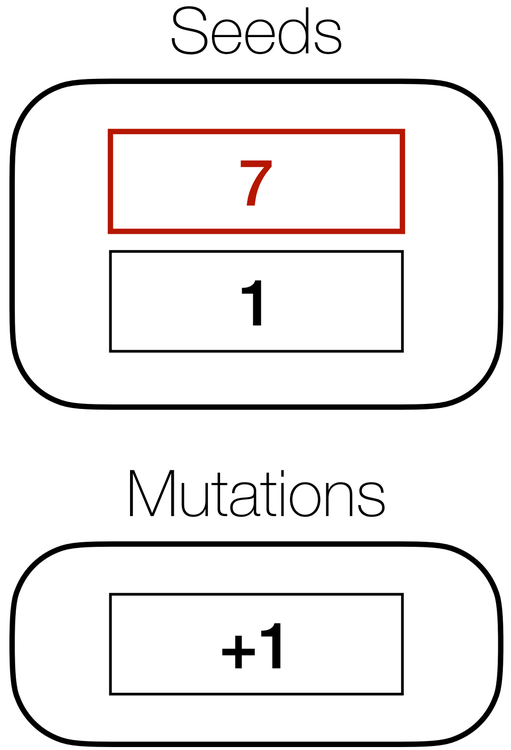


tuple = ({A1}, {B1})

Coarse path δ-diversity considers tuples of the number of covered edges.

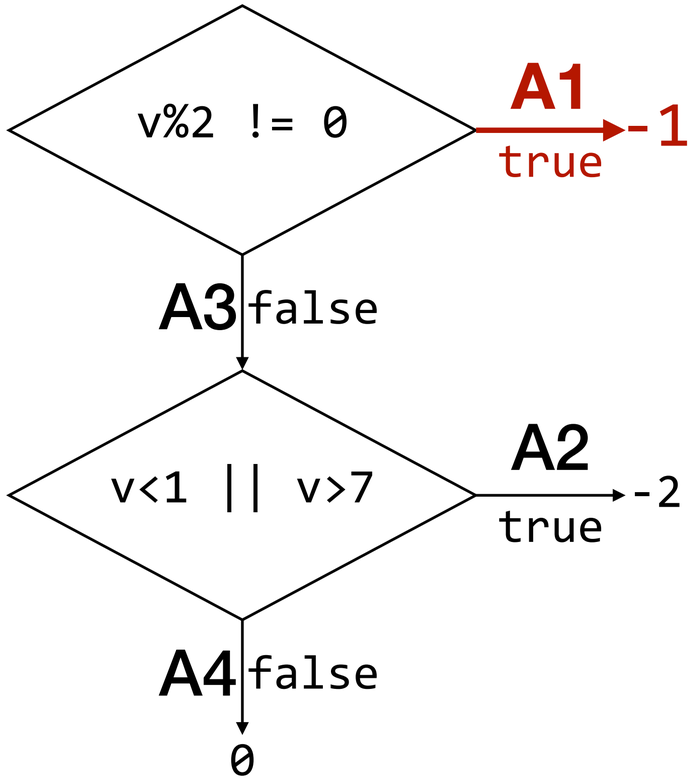


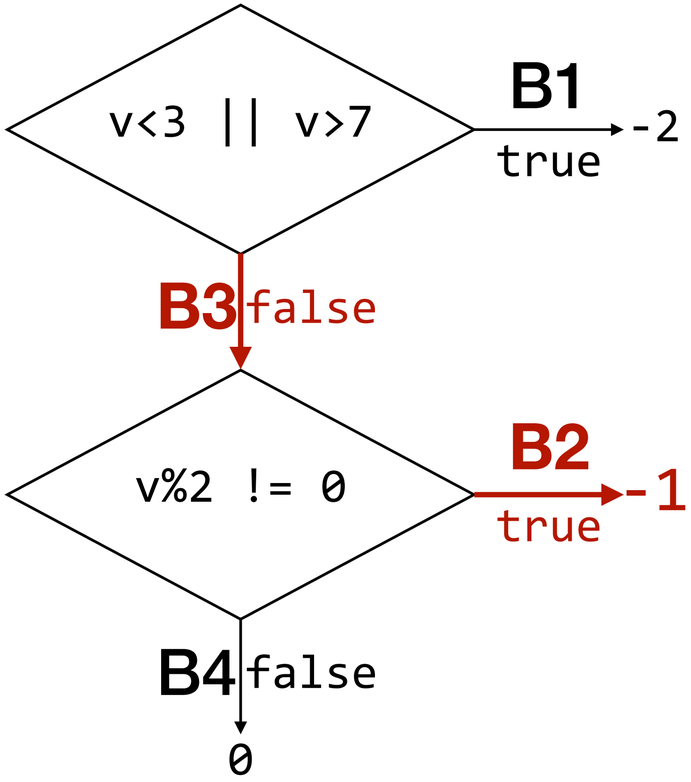


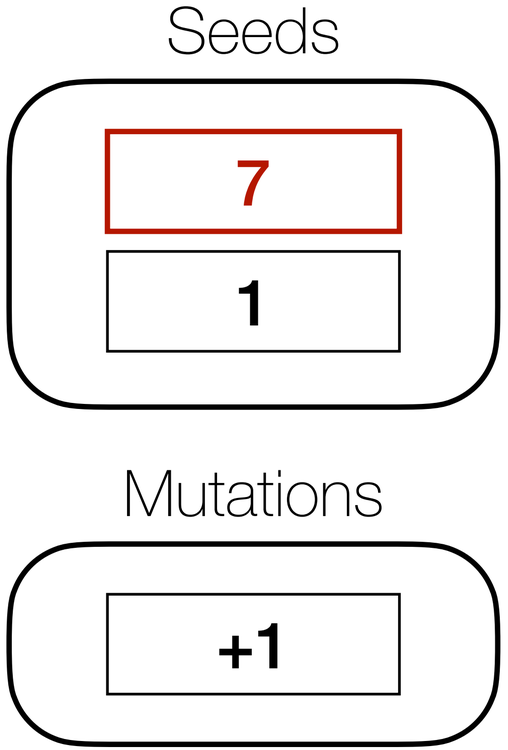


tuple = (|{A1}|, |{B2, B3}|)

Coarse path δ-diversity considers tuples of the number of covered edges.

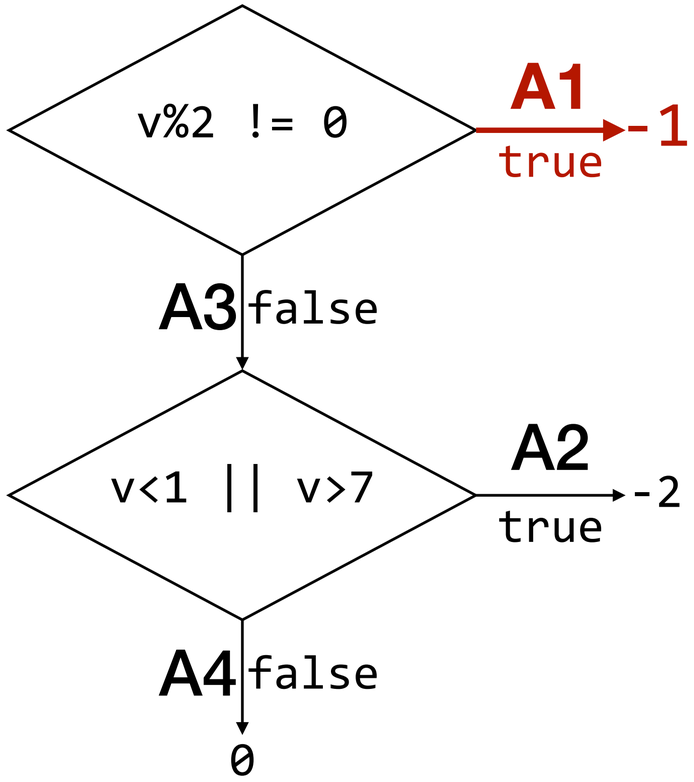


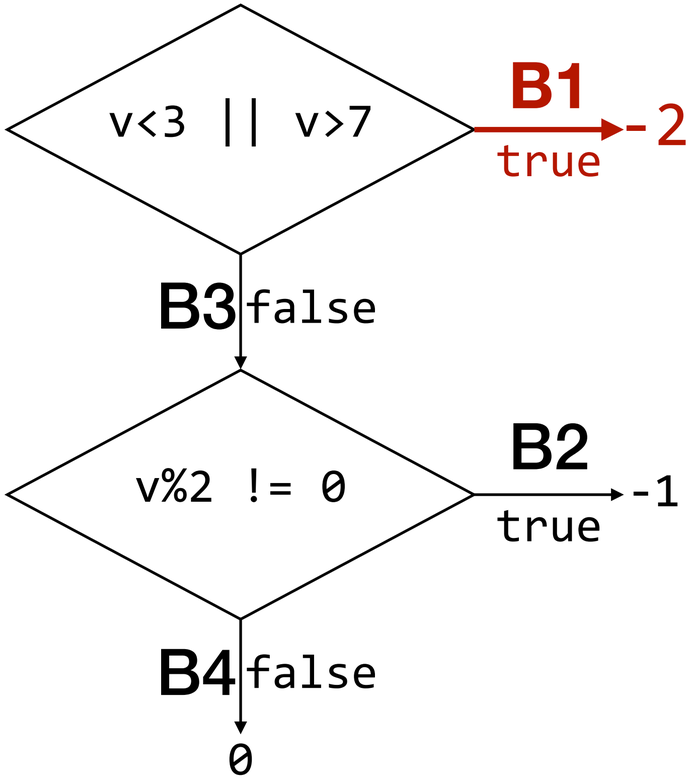


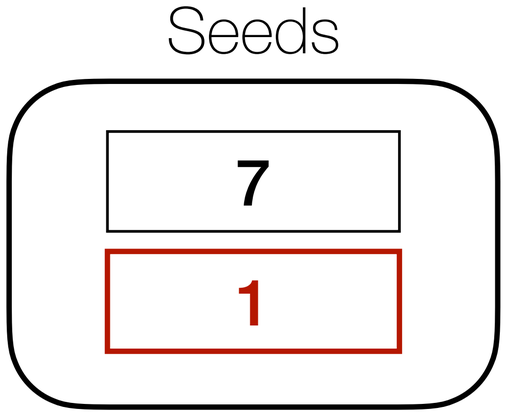


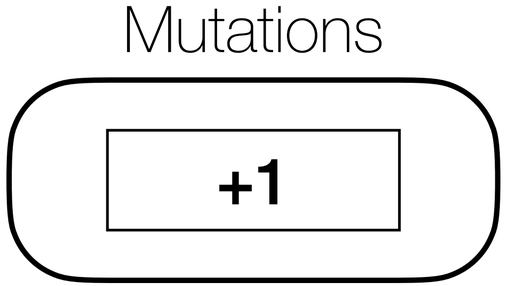
tuple = (1, 2)

Coarse path δ-diversity considers tuples of the number of covered edges.



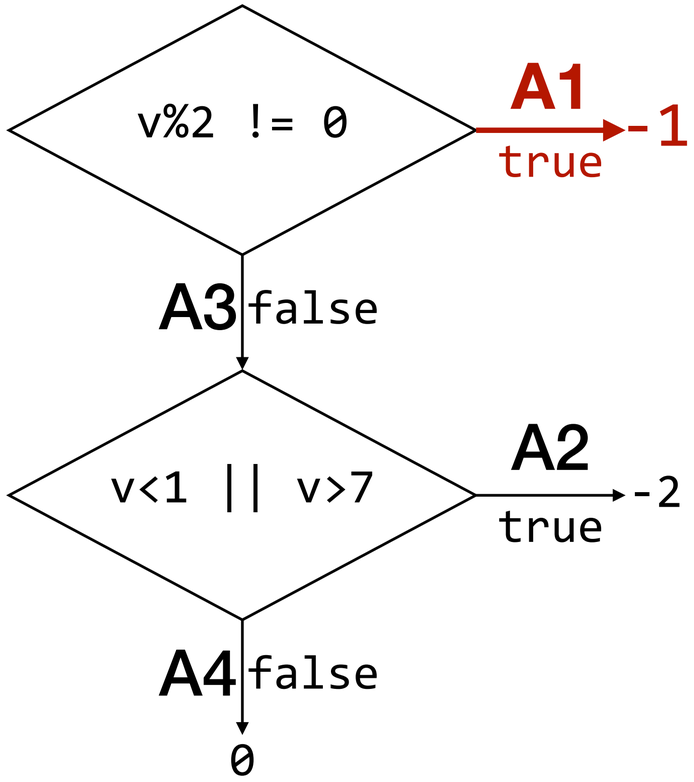


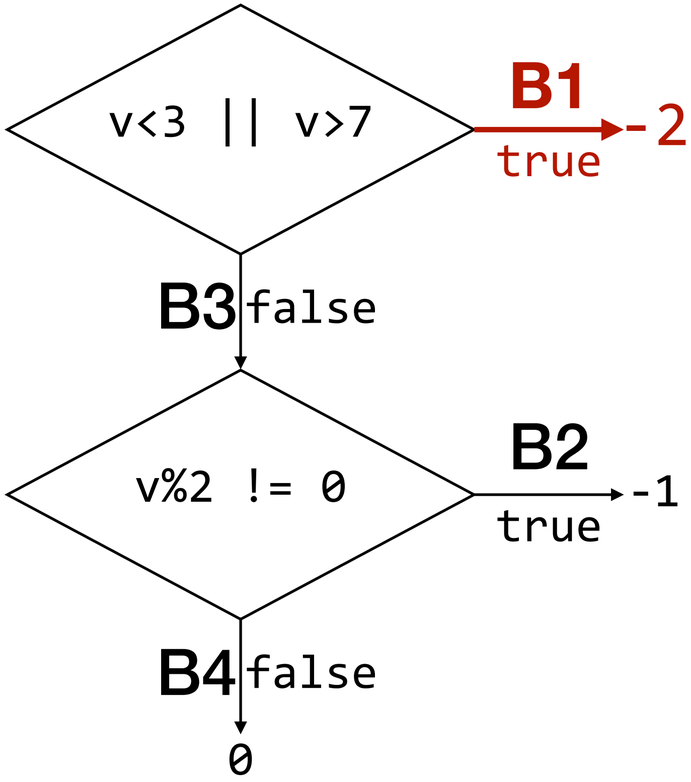


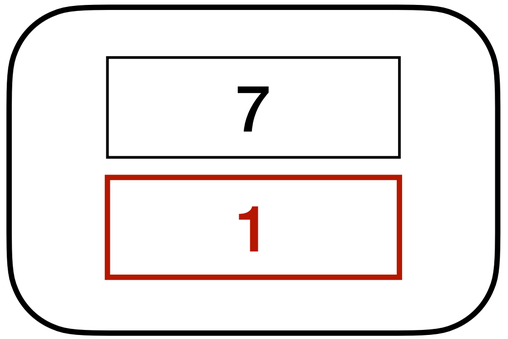


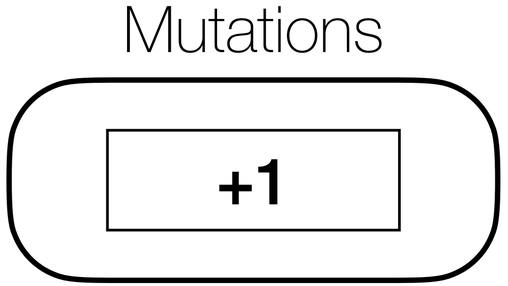
tuple = (|{A1}|, |{B1}|)

Coarse path δ-diversity considers tuples of the number of covered edges.



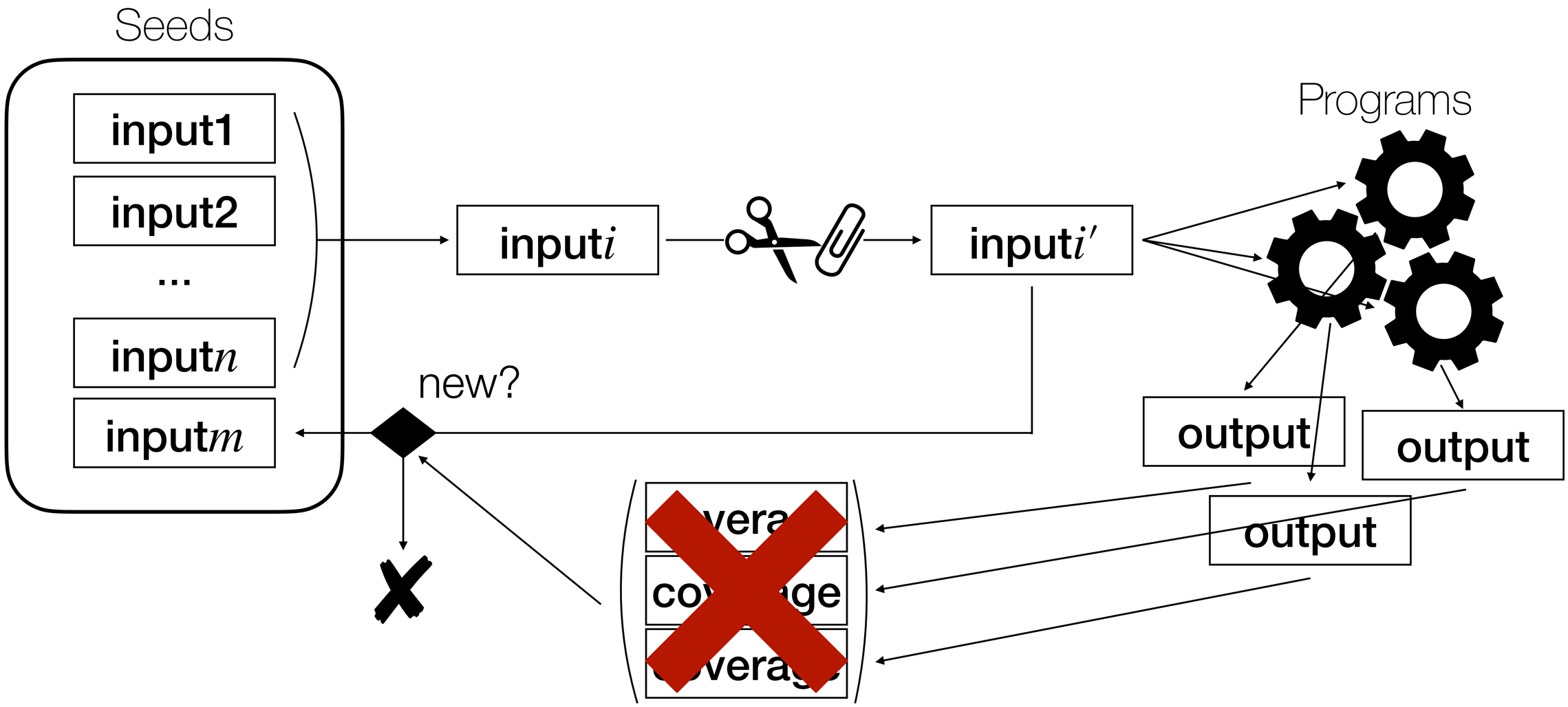




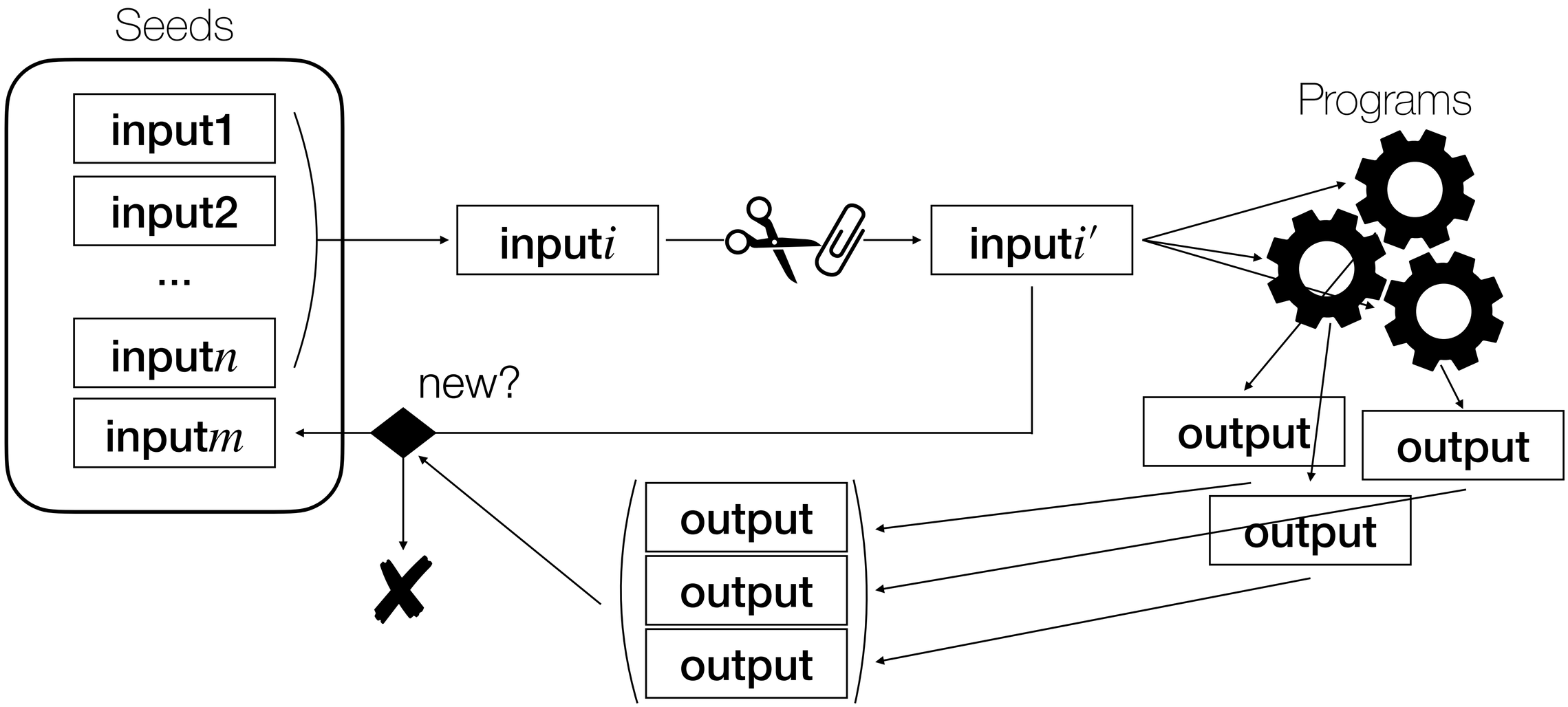


tuple = (1, 1)

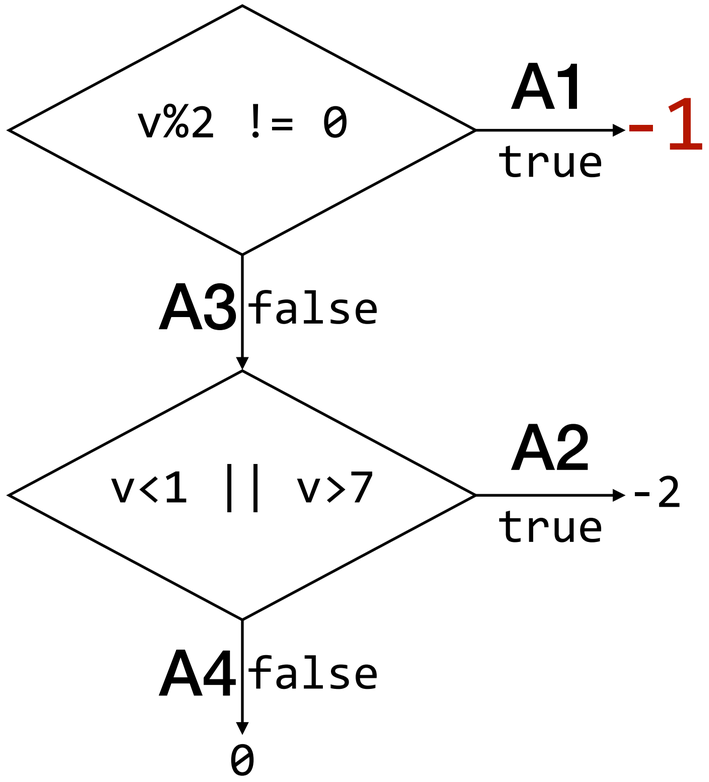
NEZHA works even in a black-box setting.

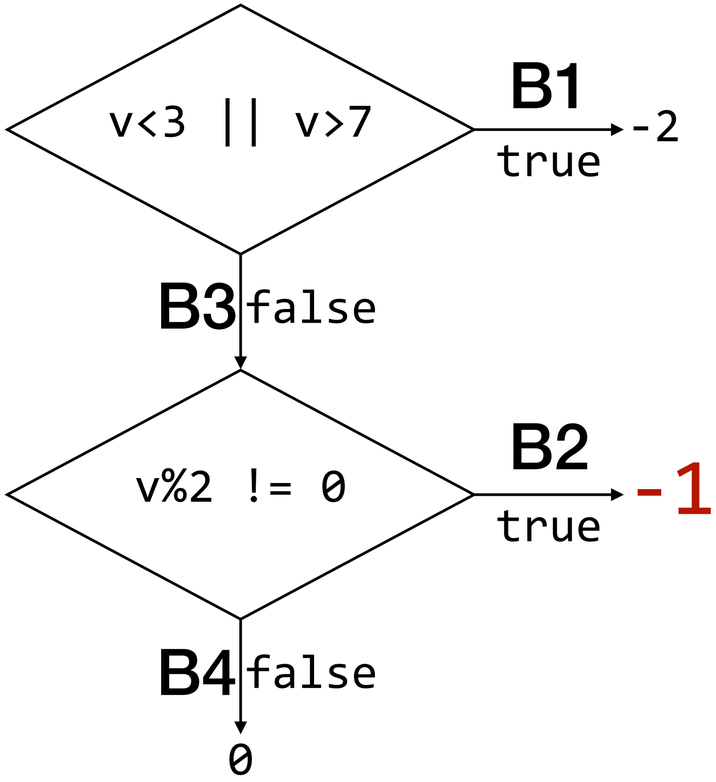


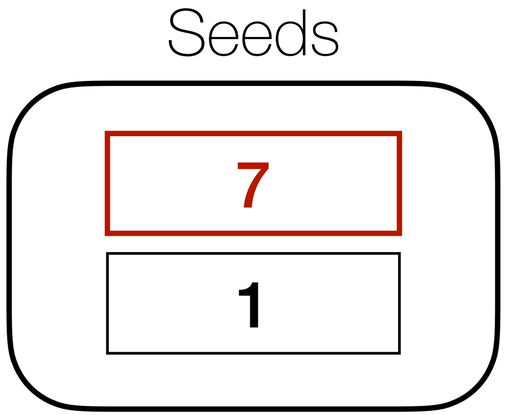
NEZHA works even in a black-box setting.

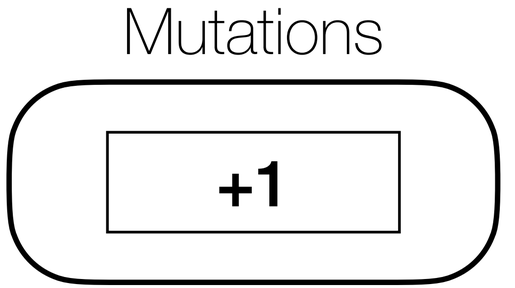


Output δ-diversity considers tuples of outputs.



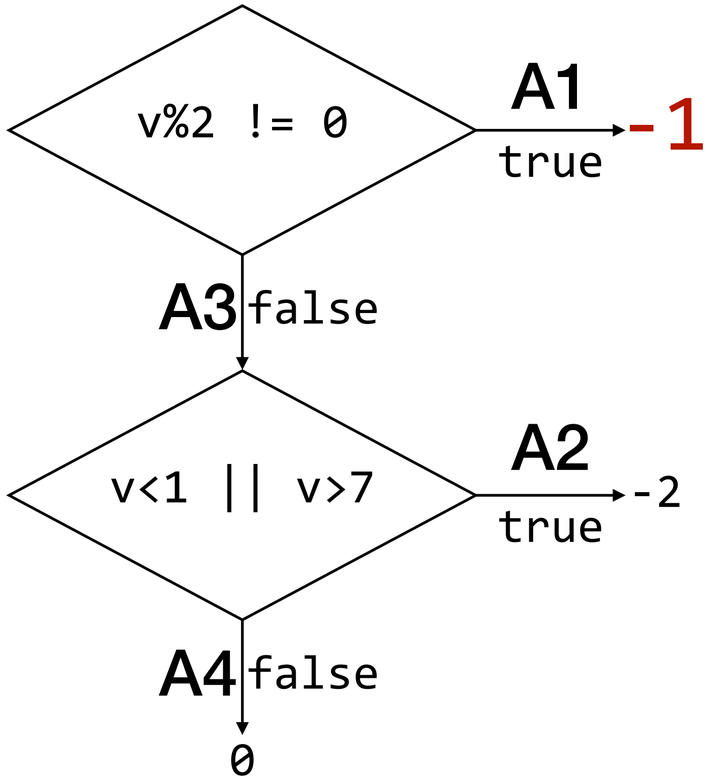


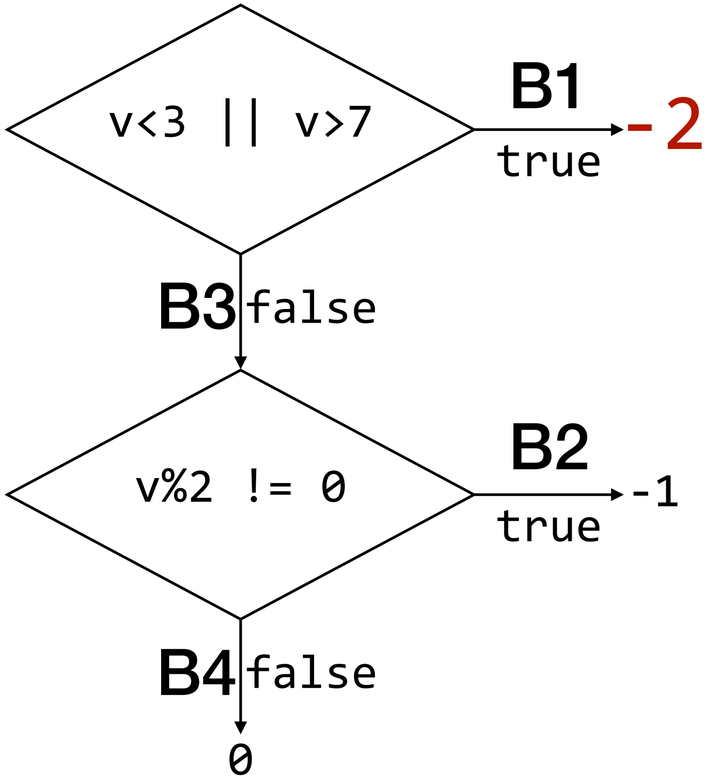


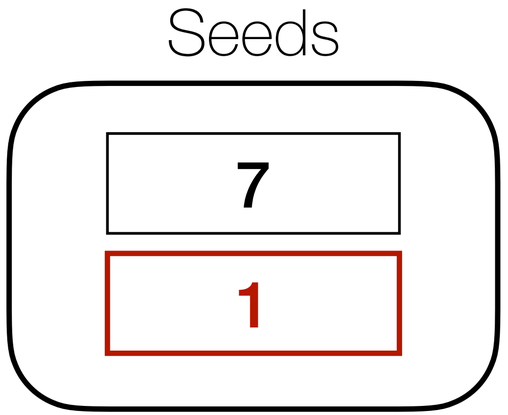


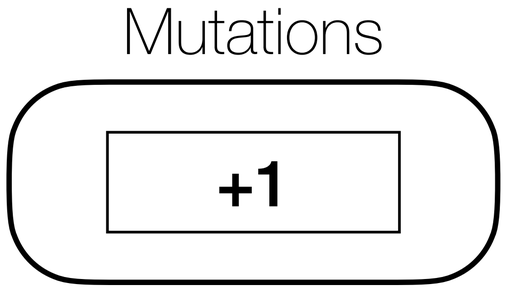
tuple = (-1, -1)

Output δ-diversity considers tuples of outputs.









tuple = (-1, -2)

Evaluation

1. Effectiveness
2. NEZHA vs Domain-specific differential testing
3. NEZHA vs Domain-independent fuzzers
4. Path δ-diversity vs Output δ-diversity

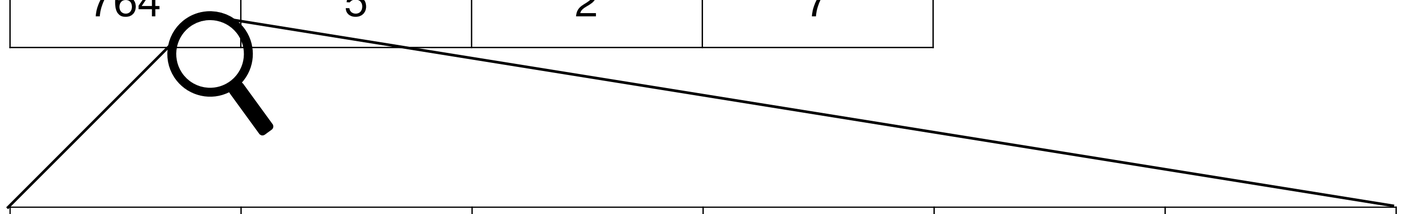
NEZHA **effectively discovers discrepancies.**

SSL/TLS	XZ	ELF	PDF
764	5	2	7

# of discrepancies

NEZHA effectively discovers discrepancies.

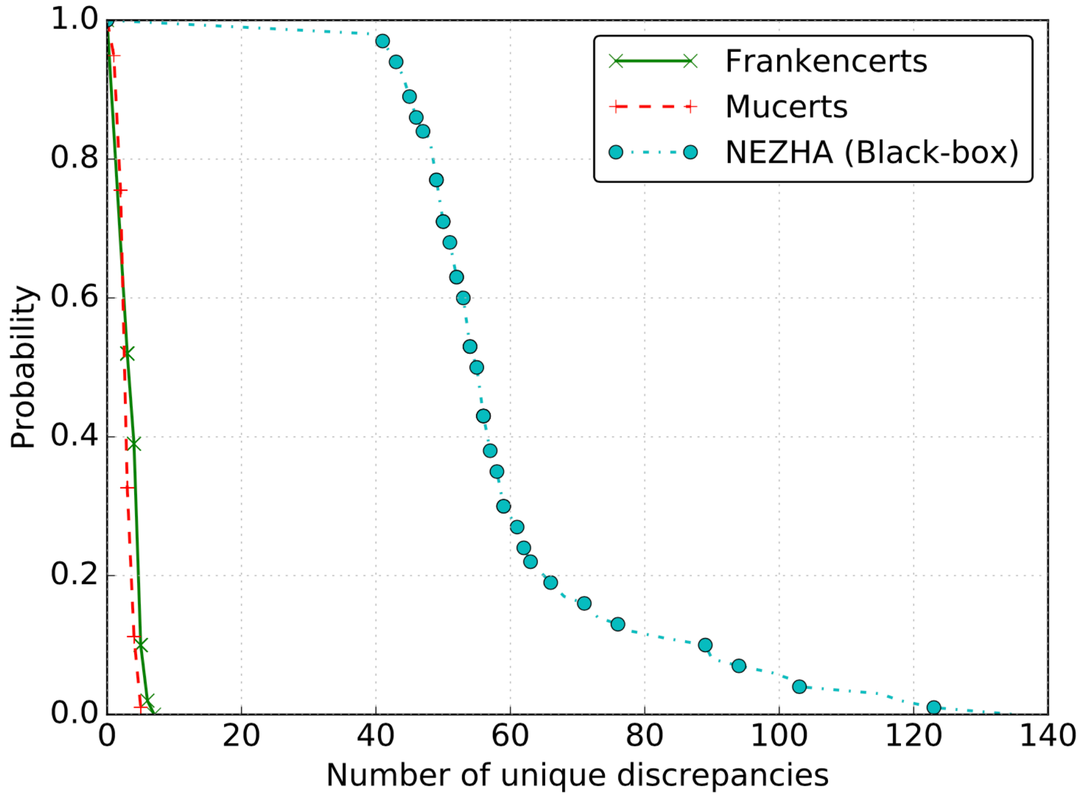
SSL/TLS	XZ	ELF	PDF
764	5	2	7



	LibreSSL	BoringSSL	wolfSSL	mbedTLS	GnuTLS
OpenSSL	10	1	8	33	25
LibreSSL		11	8	19	19
BoringSSL			8	33	25
wolfSSL				6	8
mbedTLS					31

# of discrepancies

NEZHA outperforms existing domain-specific differential testing frameworks.

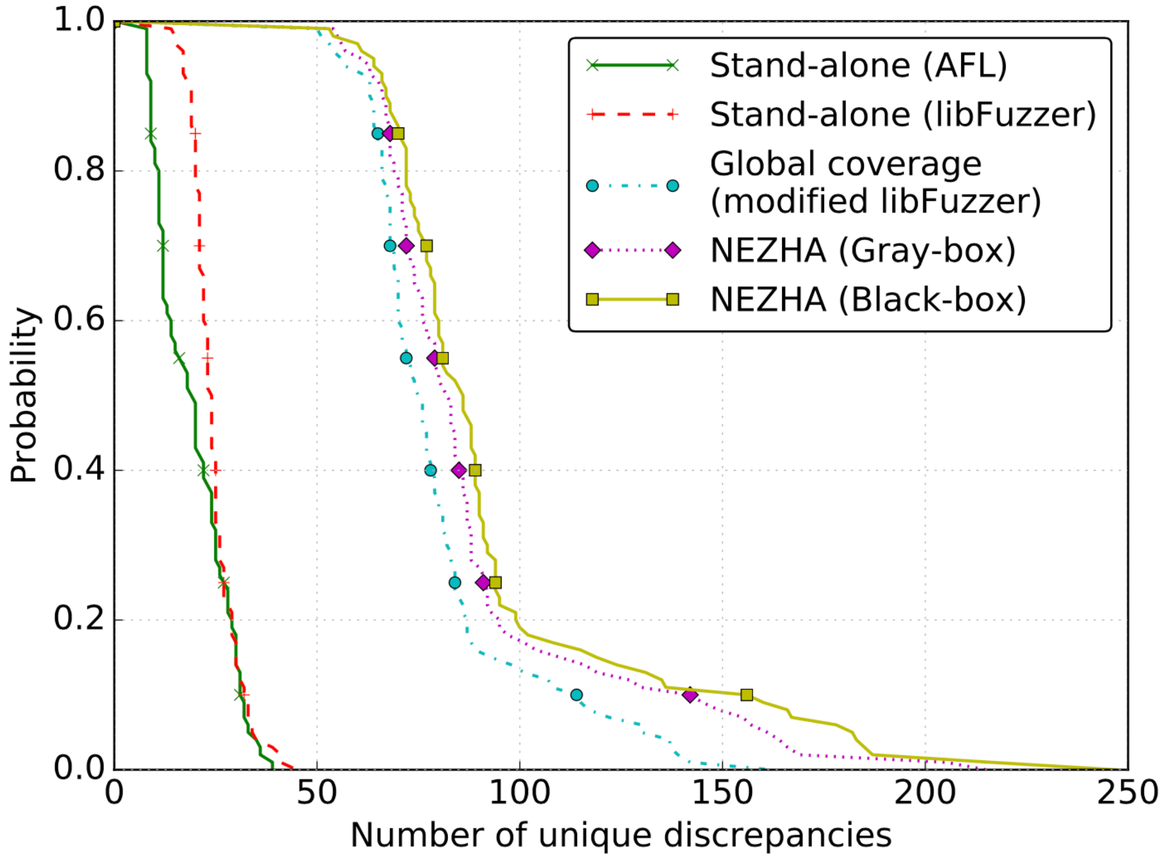


Mucerts
19

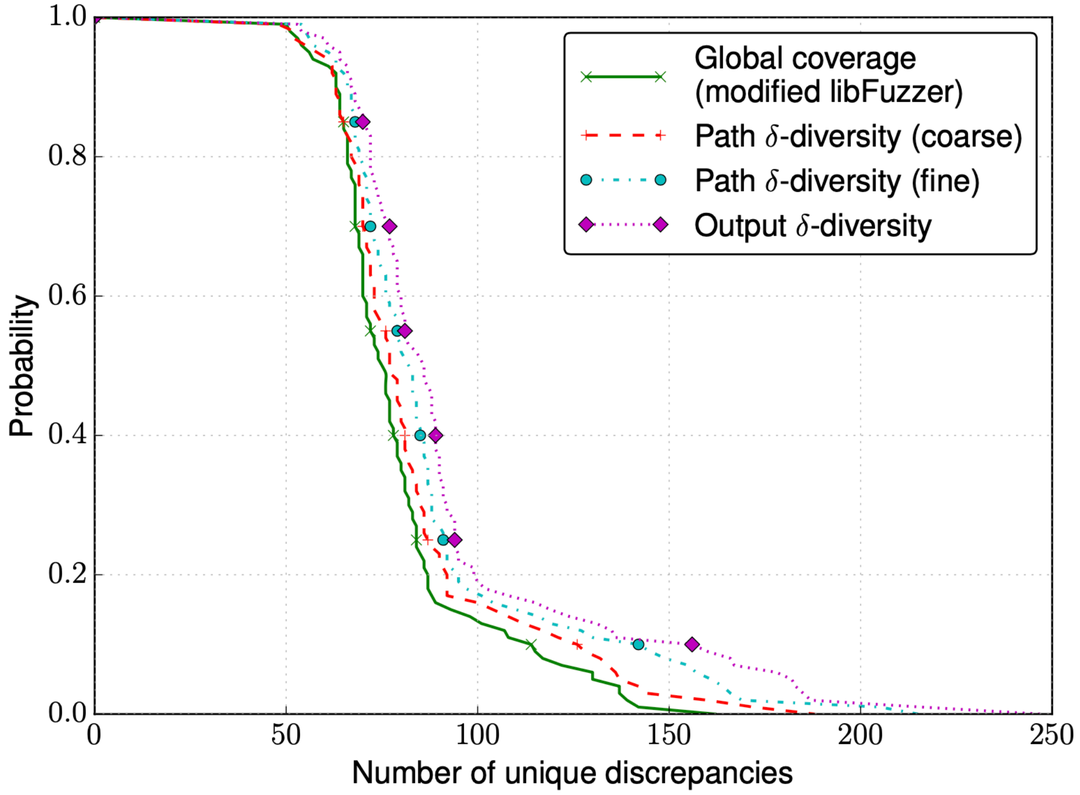
Frankencerts
10

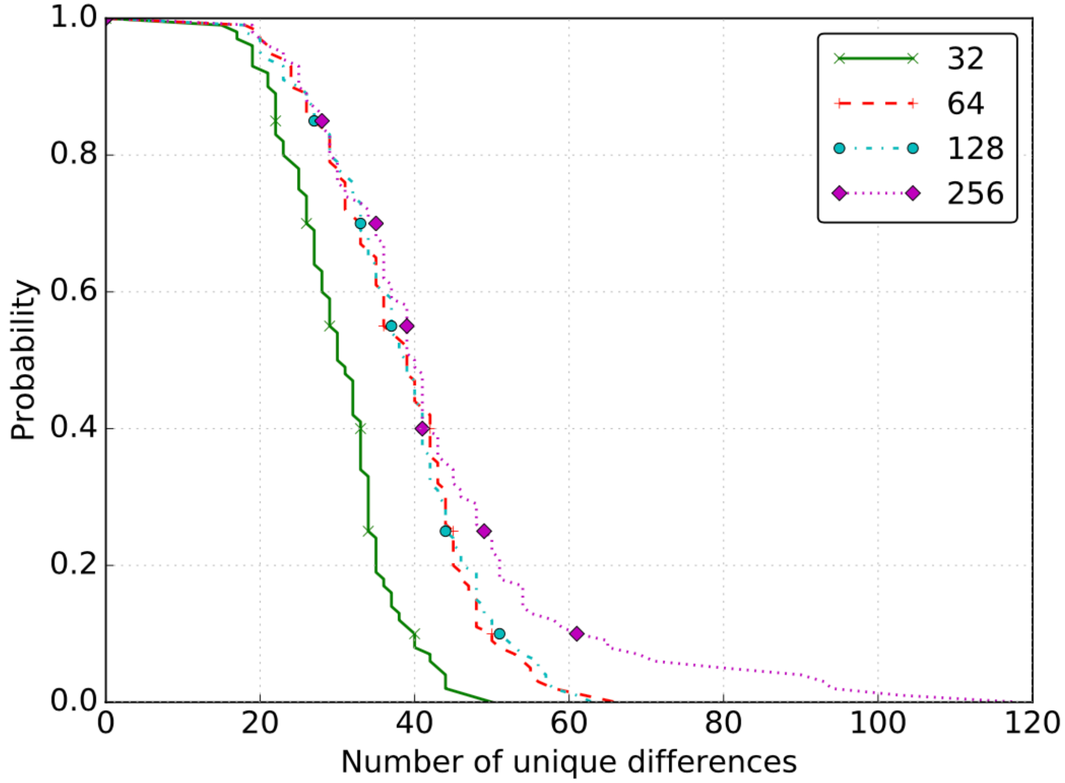
NEZHA
521

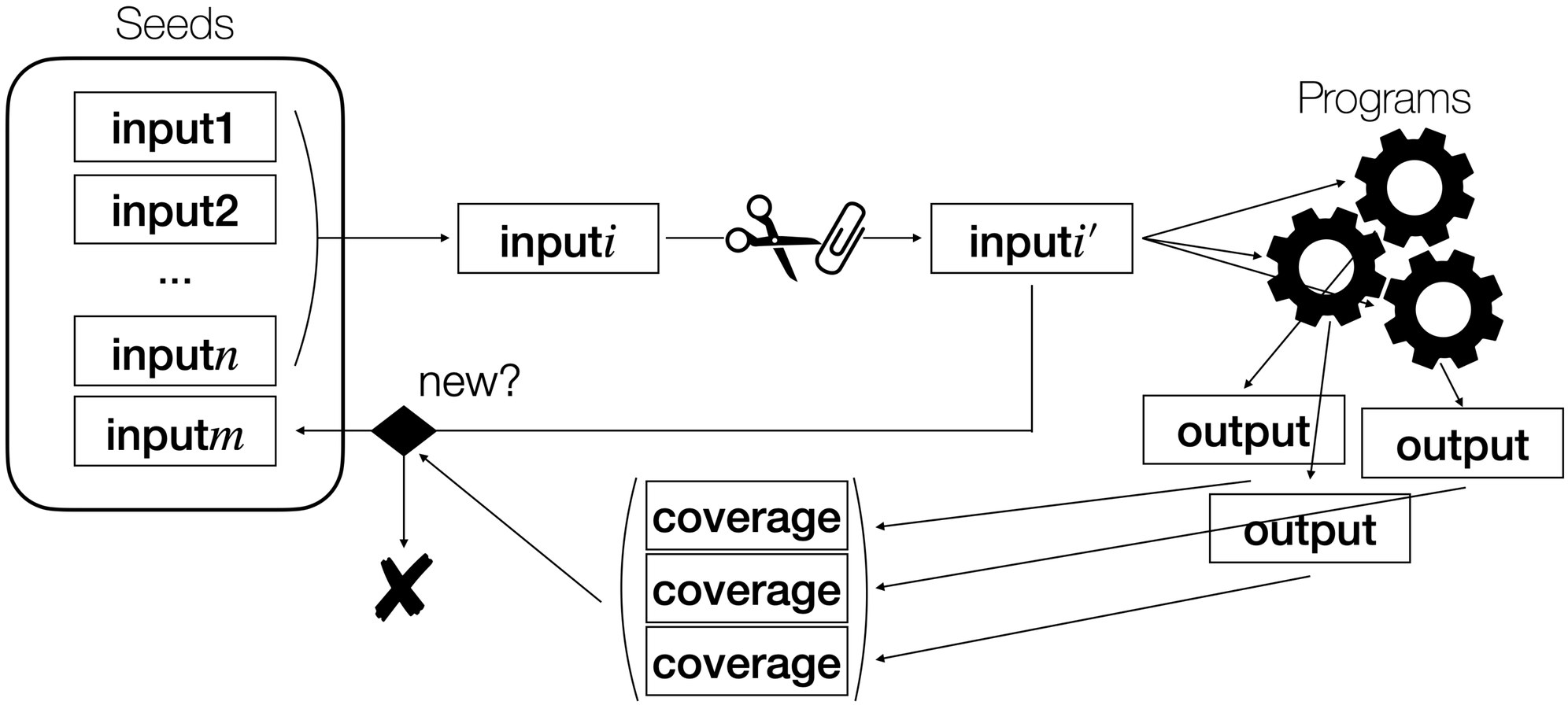
NEZHA outperforms existing domain-independent fuzzers.



Output δ-diversity is as good as path δ-diversity when outputs are fine-grained.







NEZHA

- δ-diversity
  - Fine path δ-diversity
  - Coarse path δ-diversity
  - Output δ-diversity