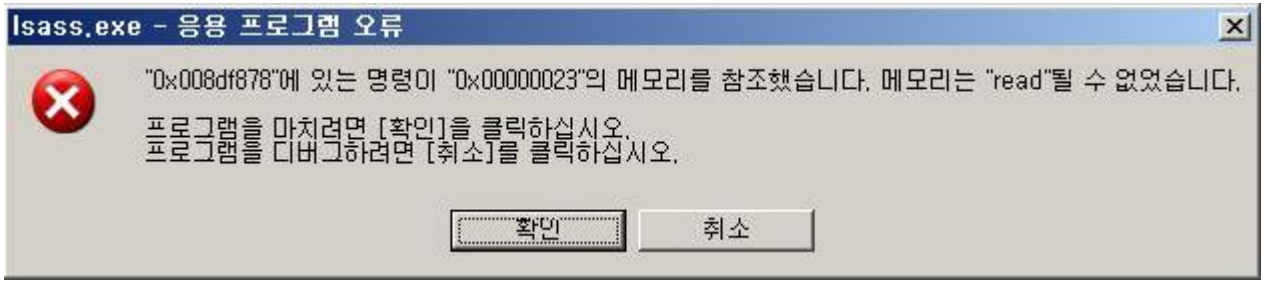
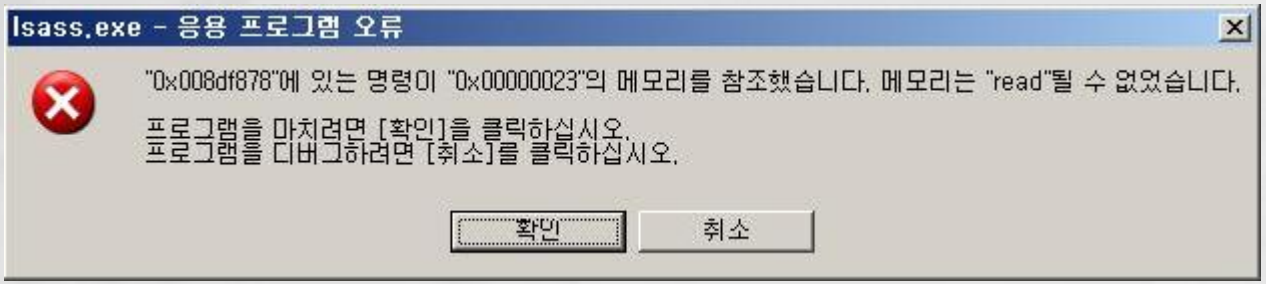
거짓 경보의 소문을 안전하게 퍼뜨리는 방법
(Alarm Clustering by Refutation)

김유일
서울대학교


lsass.exe - 응용 프로그램 오류
"0x008df878"에 있는 명령이 "0x00000023"의 메모리를 참조했습니다. 메모리는 "read"될 수 없었습니다.
프로그램을 마치려면 [확인]을 클릭하십시오.
프로그램을 디버그하려면 [취소]를 클릭하십시오.
확인
취소

버퍼 오버플로우 정적 분석



- 정적 분석으로 숨어 있는 버퍼 오버플로우 오류를 찾아보자.
- 재미있는 문제 발견/해결
  - 예: 다수의 거짓 경보(false alarm, false positive)

연구의 동기: 연결된 거짓 경보들

```
position_set grps[256];

MALLOC(grps[ngrps].elems, position, d->nleaves);
grps[ngrps].nelem = 1;
grps[ngrps].elems[0] = pos;
```

A code snippet from dfa.c in Grep 2.5.1

```
while (*optarg && *optarg >= '0' && *optarg <= '9')
    val = val * 8 + *optarg++ - '0';
```

A code snippet from ftp.c in Wu-ftpd 2.6.2

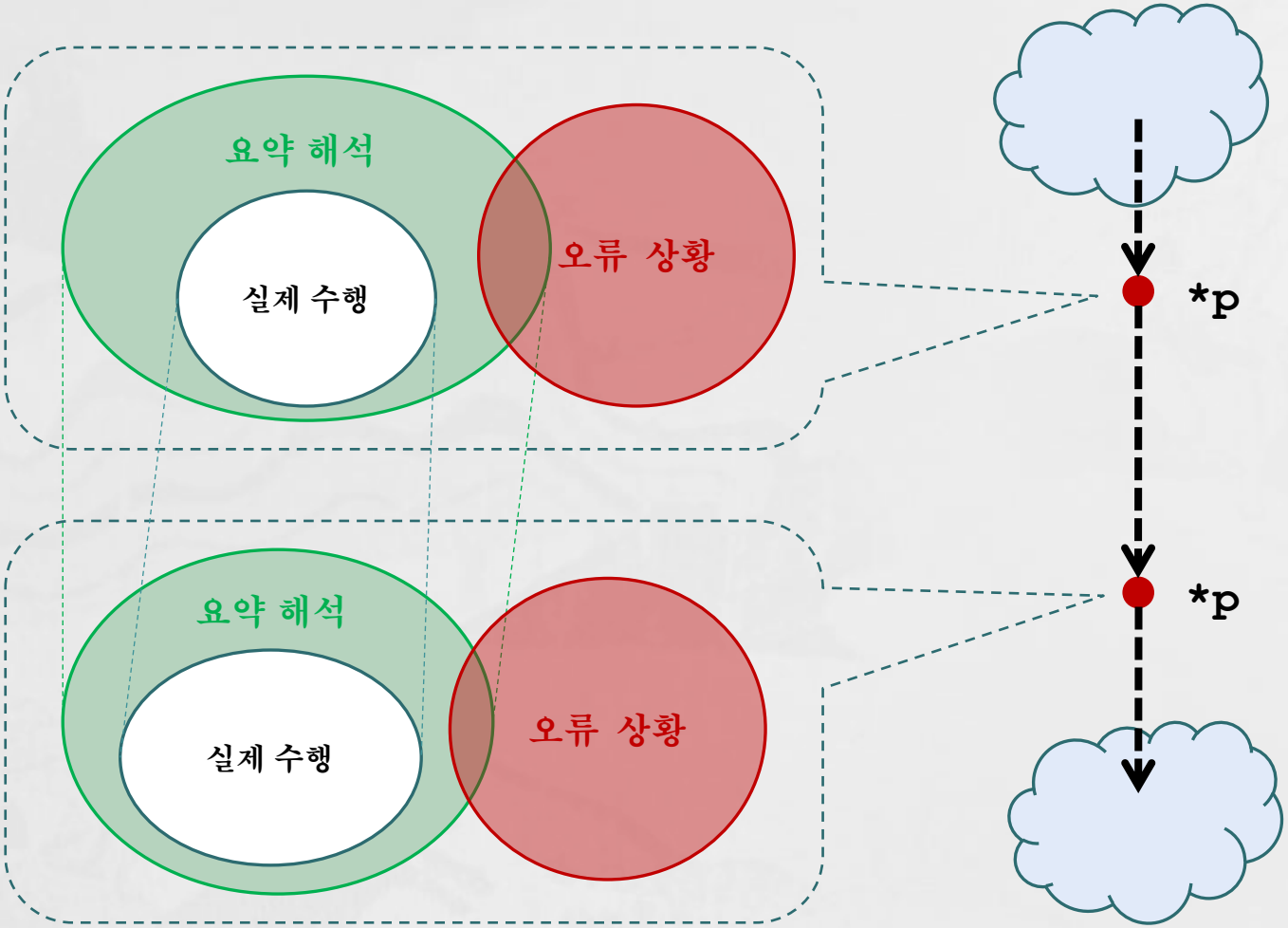
문제 정의

하나의 경보가 거짓으로 판명되었을 때,
자연스럽게 거짓임이 드러나는 또 다른 경보들을
어떻게 찾을 수 있을까?

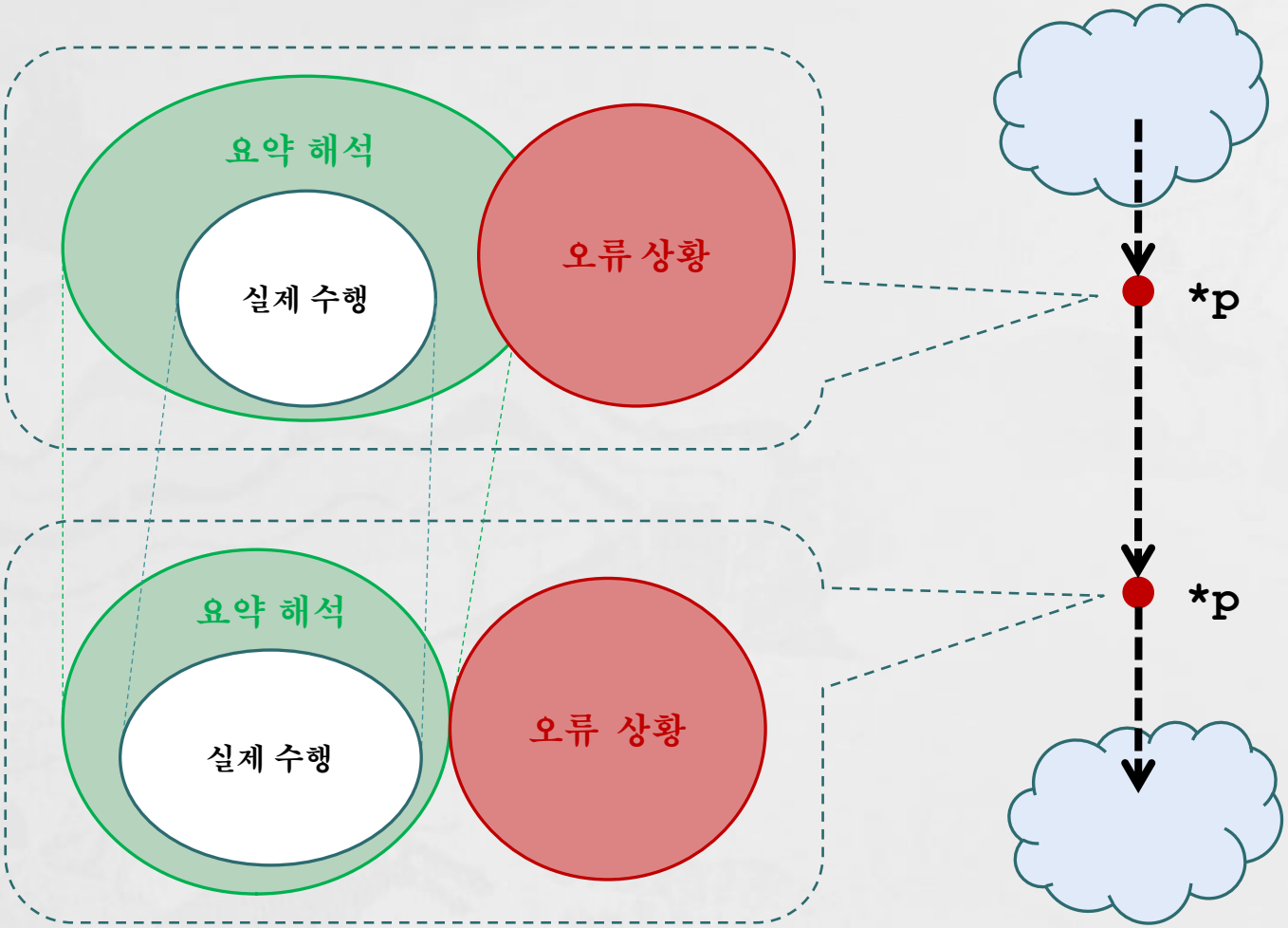
응용 가능성

- 보다 똑똑한 사용자 인터페이스
  - 사용자가 거짓 경보를 지정하면, 연결된 거짓 경보를 자동으로 찾아 제거한다.
- 보다 똑똑한 경보 보고서
  - 경보를 여러 개의 그룹으로 분류하고, 그룹별로 하나의 경보를 우선적으로 보여준다.

거짓 경보란?



거짓 경보간의 종속 관계



어려운 점

- 요약 상태로부터 오류 상태만을 “정확하게” 제외할 수 있을까?

array[i]

array[i *10+ j]

*ptr

array[i + j]

*(ptr + i)

오류 조건 제거

배열 Array의 크기는 100이고 변수 i, j의 값을 모름
{Array.size=[100,100], i=[-oo,+oo], j=[-oo,+oo]}

Array[i]

안전한 상태 집합:
0 <= i <= 99

Interval

Array[i + j]

안전한 상태 집합:
0 <= i + j <= 99

Octagon

Array[i *10 + j]

안전한 상태 집합:
0 <= i*10 + j <= 99

Polyhedron

오류 조건 제거

포인터 Ptr에 대해서 아무 것도 모름
{Ptr.offset=[-oo,+oo], Ptr.size=[0,+oo]}

*Ptr

안전한 상태 집합:
0 <= Ptr.offset < Ptr.size

Zone

*(Ptr + i)

안전한 상태 집합:
0 <= Ptr.offset + i < Ptr.size

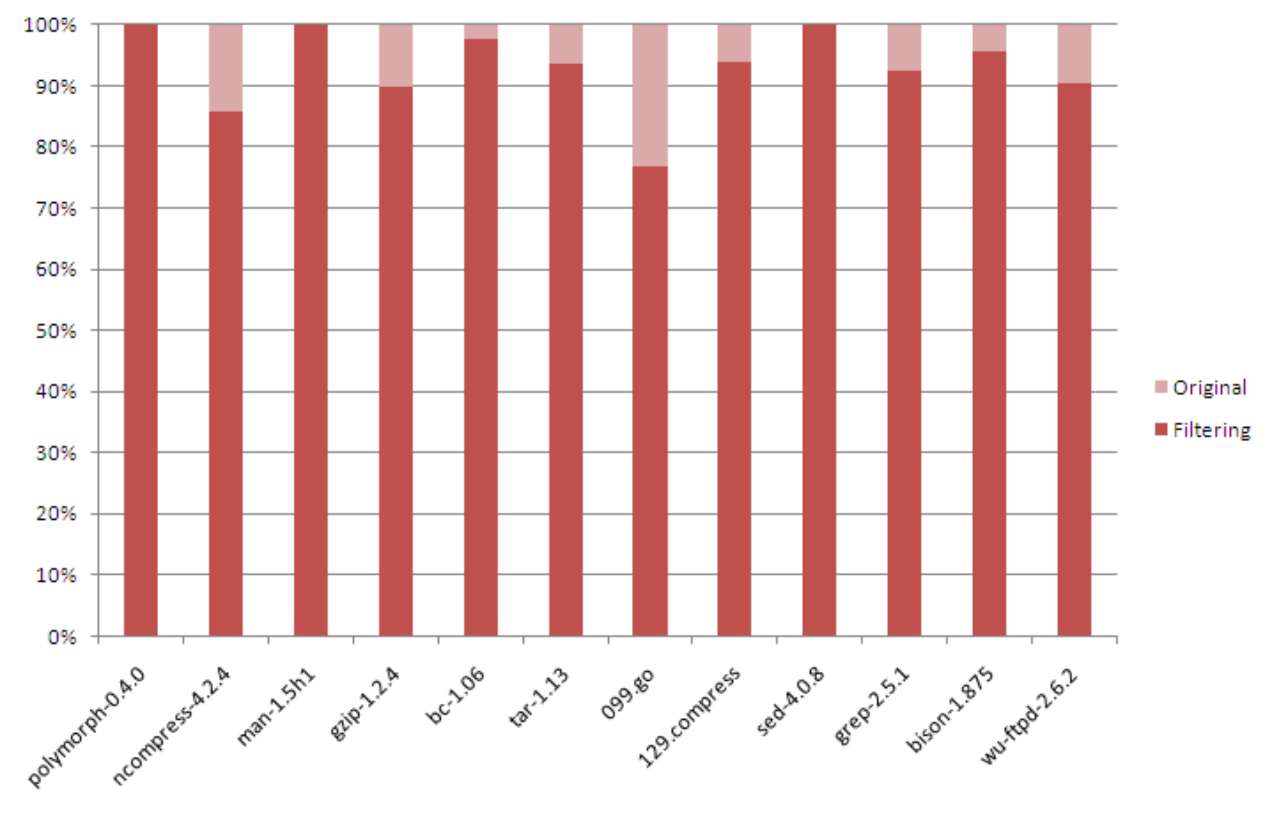
Polyhedron

예비 실험

- 연결된 거짓 경보 제거 실험: Array[i], *Ptr
- 대상 프로그램: BugBench 벤치마크 중 버퍼 오버플로우와 관련된 12개의 프로그램
- 버퍼 오버플로우 정적 분석기: Raccoon

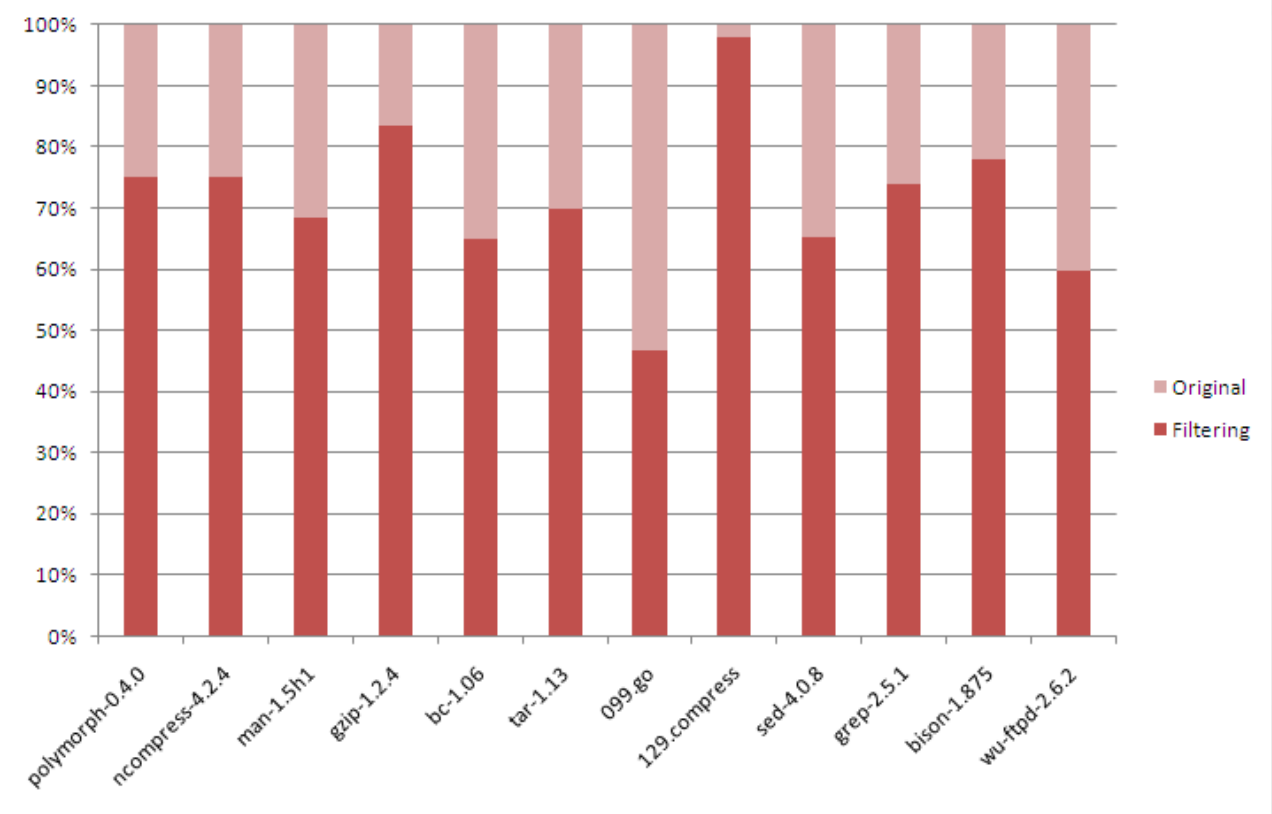
Shan Lu *et al.*, BugBench: Benchmarks for Evaluating Bug Detection Tools, In *Workshop on the Evaluation of Software Defect Detection Tools*, 2005.

예비 실험 결과: Array[i]



평균 22% 줄었음 (9,529 ▷ 7,408)

예비 실험 결과: *Ptr



평균 31% 줄었음 (11,374 ▷ 7,849)

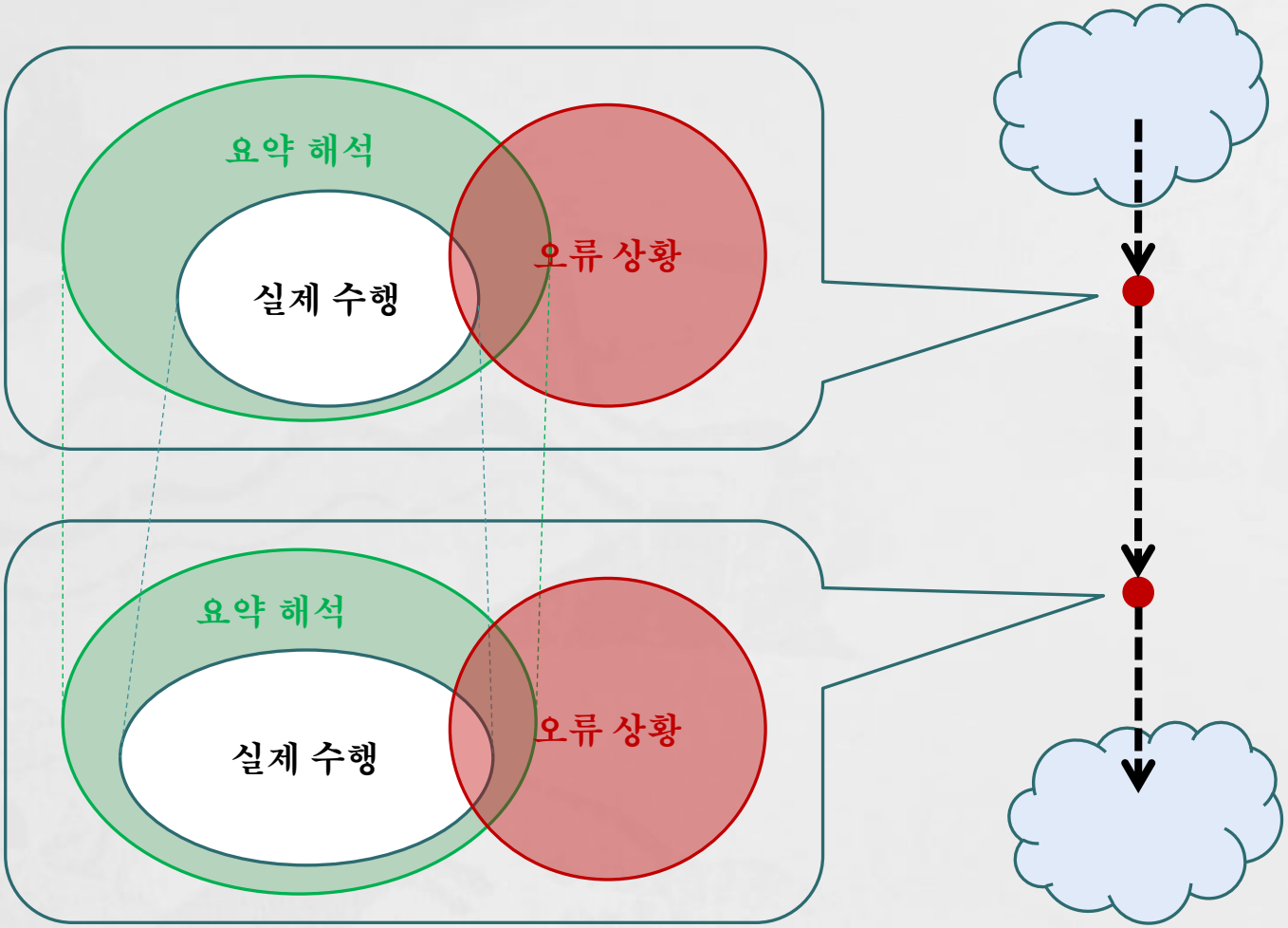
연구 계획

- 안전한 상태 집합을 식 자체로 표현
  - $(0 <= i) \wedge (i <= 99)$
  - $(0 <= i + j) \wedge (i + j <= 99)$
  - $(0 <= i * 10 + j) \wedge (i * 10 + j <= 99)$
- SMT 해결기를 이용한 Symbolic Execution
  - 요약 해석 결과를 재활용

요약 및 토의

- 경보가 "거짓 경보"라는 정보를 이용해 연결된 거짓 경보들을 자동으로 찾아내는 방법
- 경보가 "실제 오류"라는 정보를 어떻게 이용할 수 있을까?

토의: 실제 오류의 경우는?



토의: 실제 오류의 경우는?