요약해석에 기반한 경보 동일 군집 구하기

이 우 석, 김 유 일
서울대학교 프로그래밍 연구실

서론

- 정적 분석기는 많은 경보를 생성한다.
  (Sparrow 350여만 줄 휴대폰 SW 분석결과)

오류 종류	개수
Null 참조	7500
버퍼 오버런 / 언더런	765
메모리 누수	296
기타	169
총	**8730**

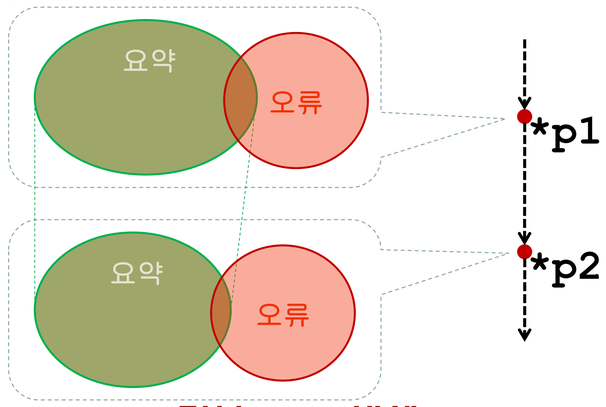
- 경보들 중 한 원인을 공유하는 것들이 많다.

```
while (1) {
    *(htab_p - 16) = m1;
    *(htab_p - 15) = m1;
    …
    *(htab_p - 1) = m1;
    htab_p -= 16;
    i -= 16L;
    if (! (i >= 0L)) {
      break;
    }
  }
  i += 16L;
```

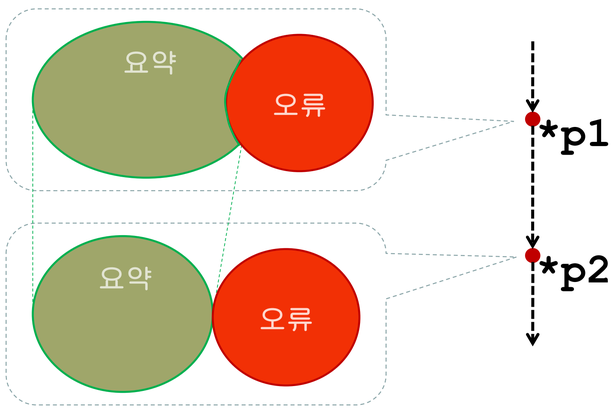
- 원인을 공유하는 경보 군집을 구하면 검사해야 할 경보 수를 줄일 수 있다.

방법 및 이론적 배경

상황.

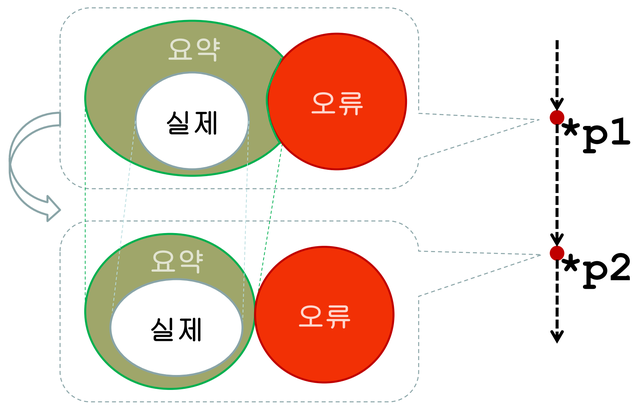


경보 1,2 발생.

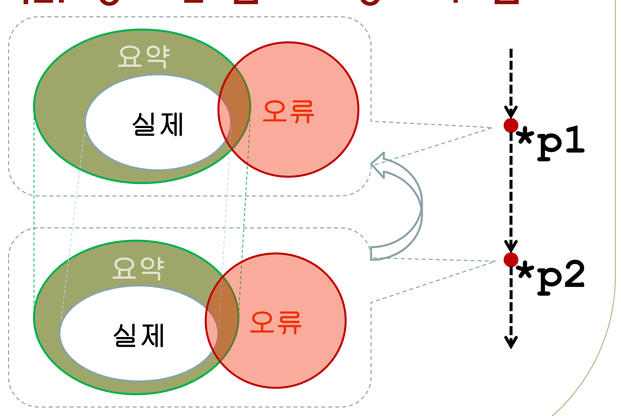


경보 1이 거짓이라는 가정하에 분석 진행 → 경보 2 사라짐.

정리1. 경보 1 거짓 ⇒ 경보 2 거짓



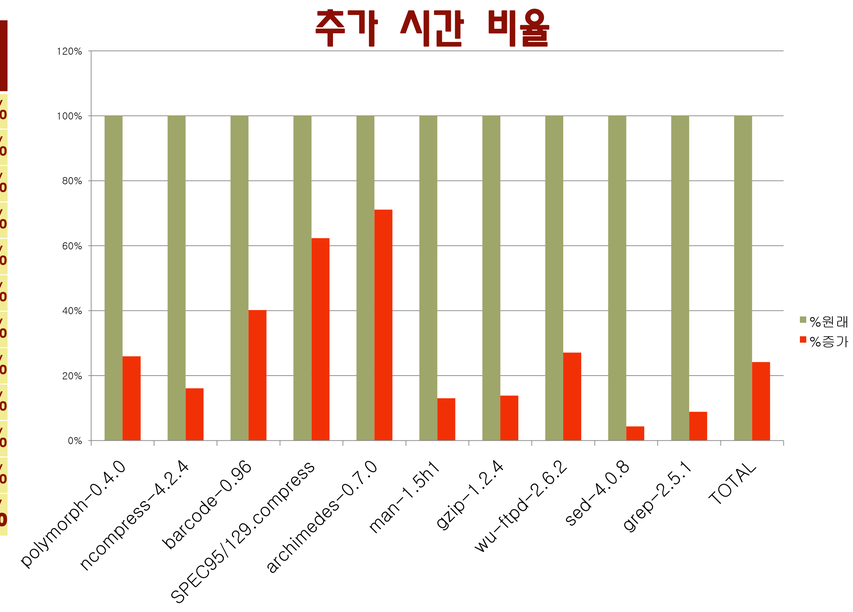
정리2. 경보 2 참 ⇒ 경보 1 참



실험 결과

- 요약해석 기반 버퍼 접근 오류 검출기 Airac에 적용

프로그램	라인 수	#경보 전	#경보 후	%감소비율	#군집	군집 크기 (평균)	군집 크기 (최대)	분석 시간	추가 시간	%증가비율
haplo-0.2.4	709	9	2	78%	2	3.5	4	0.25	0.06	24%
polymorph-0.4.0	1,357	9	3	67%	1	6.0	6	0.27	0.07	26%
ncompress-4.2.4	2,195	14	9	36%	3	1.7	2	2.67	0.43	16%
barcode-0.96	4,460	239	155	35%	39	2.7	12	15.03	6.04	40%
SPEC95/129.compress	5,585	61	41	33%	5	4.0	15	3.53	2.20	62%
archimedes-0.7.0	6,959	978	488	50%	163	3.5	12	46.40	32.99	71%
man-1.5h1	7,232	54	32	41%	10	2.5	7	110.32	14.38	13%
gzip-1.2.4	11,213	335	282	16%	40	2.4	5	66.20	9.16	14%
wu-ftpd-2.6.2	18,071	517	464	10%	29	2.6	7	2900.26	785.15	27%
sed-4.0.8	18,687	311	289	7%	6	2.2	6	426.50	18.59	4%
grep-2.5.1	20,843	30	26	13%	2	3.0	4	39.71	3.52	9%
TOTAL	96,602	2,548	1,789	**30%**	30	3.1	7.6	3611.14	872.59	**24%**



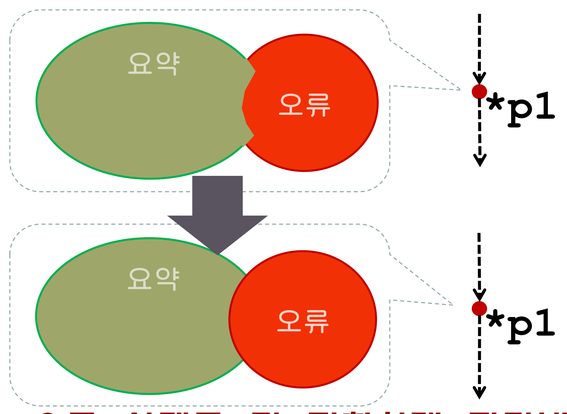
- 최대 평균 30% 경보 감소 효과. 추가 소요 시간은 평균 기존의 24%
- 하나의 경보가 거짓일 때 동시에 거짓이 되는 경보 수(군집 크기)는 평균 3개

결론

- 기존에 없던 새로운 방식
- (요약해석에 기반한) 모든 정적 분석에 적용가능
- 모든 과정이 자동
- 통계적이지 않은 논리적인 방법이므로 사용자가 결과를 확신할 수 있다.

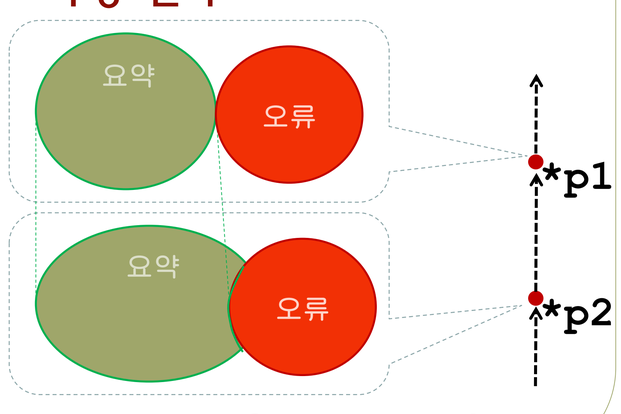
연구 계획

- 관계 분석



오류 상태를 더 정확하게 잘라냄

- 후방 분석



더 많은 관계를 유추

Programming Research Laboratory
http://ropas.snu.ac.kr

ROSAEC center
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 KOSEF ERC