SMT Solver를 활용한 알람 클러스터링

강동옥 이우석 오학주 이광근교수님
서울대학교 프로그래밍연구실

소개

SMT 알람 클러스터링?

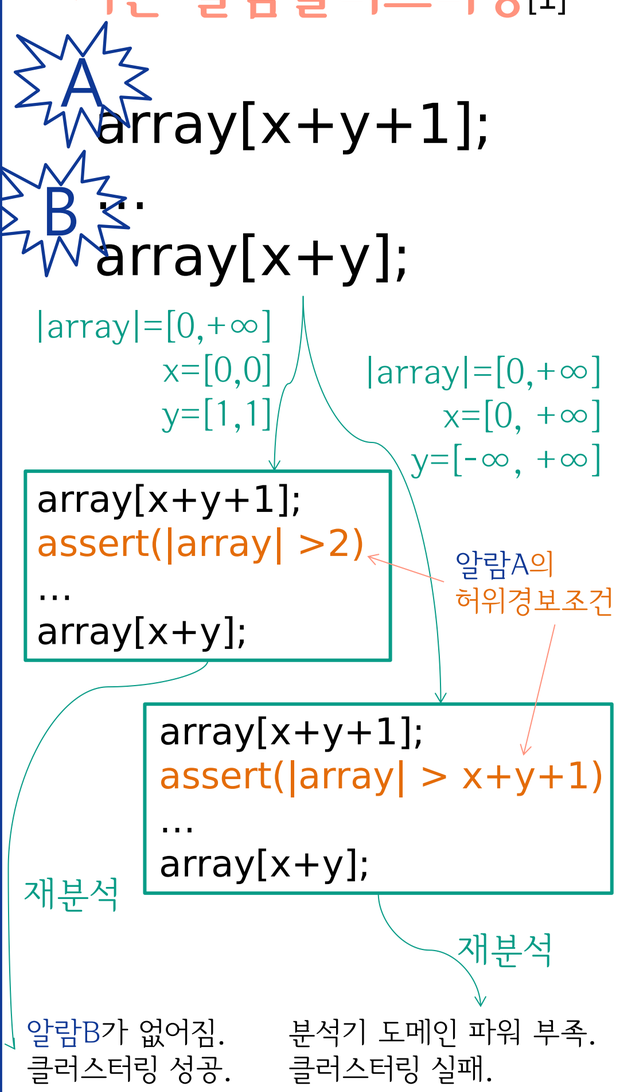
정적분석 결과로 나온 여러 알람에 대해, 대표알람만 검토해도 되도록 알람묶음을 만들어 준다.
이때 SMT Solver로 두 알람의 논리적 관계를 추론한다.

의미는?

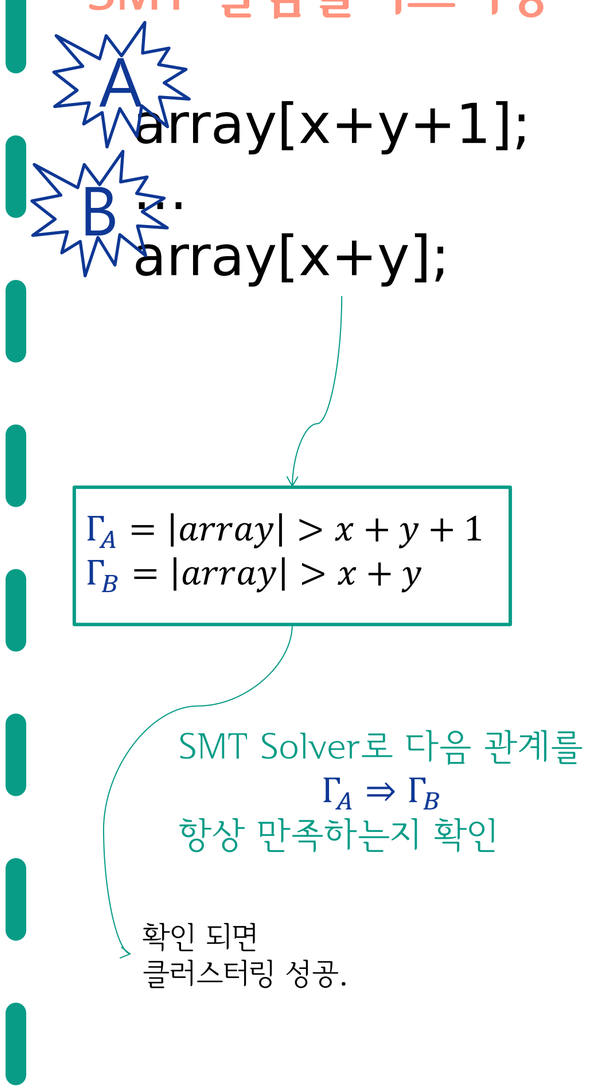
SMT Solver의 능력에 기대어 사용자가 확인해야 할 허위알람 수를 줄일 수 있다.

기존 알람 클러스터링과의 비교

기존 알람클러스터링[1]



SMT 알람클러스터링



알람클러스터링 파워
기존 알람클러스터링: 분석 요약도메인의 파워가 강력할 수록 알람이 많이 묶인다.
SMT 알람클러스터링: Solver의 표현 가능한 식이 많을수록 알람이 많이 묶인다.

각각의 장점
기존 알람클러스터링: 분석기의 고정점 계산 능력을 잘 활용하는데 강점 (ex: 다른 함수간 알람쌍)
SMT 알람클러스터링: 관계정보를 잘 활용하는데 강점 (ex: 3변수 이상이 관여된 알람쌍)

알고리즘

Step1. 비교할 알람 쌍 고르기

모든 쌍 비교는 비용이 크므로, 같은 함수 내에서 발생한 알람으로 클러스터링을 시도한다.

Step2. 허위경보 조건 논리식으로 표현

프로그램 변수를 논리식 변수로 활용하기 위해 SSA로 변환한다.
이를 토대로 허위경보조건 Γ_A, Γ_B을 각각 생성한다.

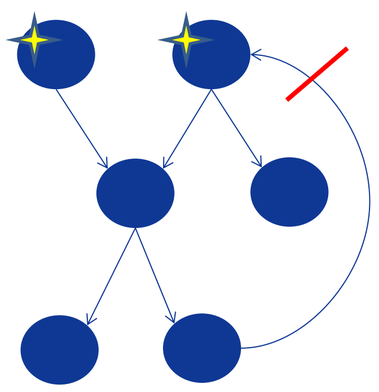
Step3. 두 알람 사이의 논리적 관계 추론

SMT Solver로 다음을 확인하여 항상 $\Gamma_A \Rightarrow \Gamma_B$ 인지 확인한다.

$$UNSAT(\neg(\Gamma_A \Rightarrow \Gamma_B))$$

Step4. 대표 알람 뽑기

클러스터링 관계에 있는 알람들로 그래프를 그린다.



선행노드가 없는 노드를 대표알람 삼아서 알람묶음을 만든다.

cycle이 존재하면 클러스터링이 많이 되도록 잘라준다.

아직 끝나지 않은 여정

Heap location SSA 만들기

현재: address taken없는 지역변수만 SSA후 논리식 생성.
(전역변수,location, address taken변수 무시)
계획: 두 알람 사이에 function call이 없는 경우 SSA로 만들기
기대: 관찰에 기반하여 긍정적으로 더 많은 클러스터링 기대.

불변식 분석

아래는 bc-1.06에서 발견한 예제이다.

$\Gamma_B = |ptr0| > ptr0.offset$
$\Gamma_A = |ptr1| > ptr1.offset$
$I = (|ptr1| = |ptr2|)$
$\wedge\ (ptr1.offset = ptr0.offset + 1)$

I가 항상 true이므로
$UNSAT(\neg(I \Rightarrow \Gamma_A \Rightarrow \Gamma_B))$를 확인하면
알람 A,B는 클러스터링 된다.

결과

기존 알람 클러스터링으로 줄인 알람 개수에 대해 **28.1%**의 알람이 추가적으로 감소

프로그램	코드크기(LOC)	분석시간(sec)	SMT클러스터링시간(sec)	기존클러스터링 후 알람 (#)	SMT클러스터링 후 알람(#)	감소비율(%)
nlkain-1.3	831	2	2	79	68	13.92
xlreader-0.9.0	2705	3	2	145	60	58.62
pk7	3813	8	<0.1	13	12	7.69
archimedes-1.1.0	6882	133	46	313	242	22.68
smail-3.2	11251	95	<0.1	5	3	40.00
bc-1.06	15846	733	24	286	212	25.87

[1] Woosuk Lee, Wonchan Lee, and Kwangkeun Yi, Sound Non-statistical Clustering of Static Analysis Alarms, VMCAI 2012

SEOUL NATIONAL UNIVERSITY

Programming Research Laboratory