

---

# PLDI 2014 여행기

에딘버러, 스코틀랜드

2014년 6월 8일 - 2014년 6월 15일

서울대학교 허기홍

---



아름다운 자연과 고풍스런 도시가 어우러진 스코틀랜드 에딘버러에서 열린 PLDI 2014 학회에 다녀왔다. 프로그래밍 언어 분야의 최고 학회답게 큰 규모와 훌륭한 참가자들에 둘러싸여 흥분된 나날들이었다. 또한, 우리가 쓴 논문이 채택되는 경사도 함께 안고 갔다. 학회는 근사한 회의장 같은 건물에서 열렸다. 두 세션이 병렬로 진행되고 프로그래밍 언어분야의 다양한 주제에서 논문이 나왔다. 2년전 PLDI에 참가했을때와 다른 특징은 아무래도 최근 떠오르는 확률적 프로그래밍 관련 세션이 생겼다는 점이다. 큰 물결이 되어가는 듯 하다.

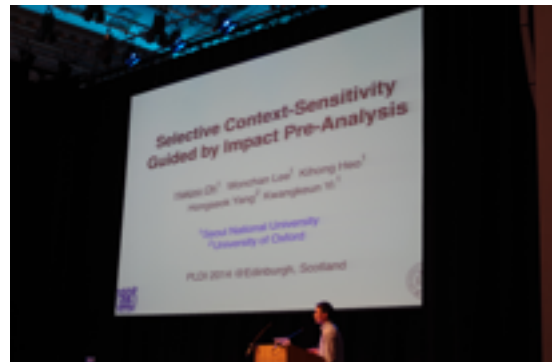


## 논문 발표

### Selective Context-sensitivity Guided by Impact Pre-Analysis

오학주, 이원찬, 허기홍, 양홍석, 이광근

지난 2012년에 이어 이번에도 저자로서 기쁜 마음으로 참가하게 된 논문이다. 이번 논문은 비싼 분석을 꼭 필요한 곳에 선별적으로 적용함으로써 정확도와 확장성을 함께 높이는 방법이다. 예를 들면, 문맥 구분 분석이나 관계 분석 같이 정확하지만 비싼 분석을 선별적으로 적용하는 것이다. 핵심 기술은 정확도 상승 효과를 가늠하는 전분석 (Impact Pre-Analysis) 이다. 전분석은, 문맥이나 관계처럼 우리가 선별적으로 적용하고 싶은 요소를 전혀 요약하지 않음으로서 본분석에서 필요한 요약 정도를 찾아낸다. 대신, 나머지 의미 요소는 과감하게 요약하여 계산이 빨리 끝날 수 있도록 만든 것이다. 이 기술을 기반으로 문맥 구분 분석과 관계 분석을 우리 연구실의 정적분석기인 SPARROW에 적용한 결과 눈에 뛴 만큼 분석기가 정확해졌다.



학주형이 발표를 했는데 우리가 지금까지 했던 일의 큰 흐름이 잘 전달된 것 같다. 발표 내용을 준비할 때 어깨너머로 배우고, 토론하며 수정하는 과정에서 많은 생각을 하였다. 요약하면, 발표에서 ‘기술’을 강조할 것인가 ‘철학’을 강조할 것인가하는 고민이었다. 결과적으로, 우리가 만든 기술이 어떻게 동작하는지 그 내용보다는, 우리가 어떤 문제 의식을 갖고 무슨 문제를 푼 것인지에 무게를 두었다. 발표장에는 제대로된 분석기를 만들어본 연구자들이 많았는데 발표후에 우리 논문에서 쓴 기술에 대해 관심을 보인 것을 보니 성공한 듯 싶다.

---

많은 것을 배우고 즐겁게 연구했던 기억이 남는다. 지금 다 정리하고 난 뒤에는 당연한 내용이  
지만 잘 이해하지 못하던 때에는 많이 혼란스러웠다. 주로 선별적 관계 분석을 연구했는데 하루  
에 한 번씩 학주형과 정의를 이렇게 저렇게 고치고 논의를 거듭했다. 그러던 중 등교 셔틀버스 안  
에서 불현듯 떠오른 정의를 잊지 않고 연구실로 가져오던 그날 아침이 생생하다. 열심히 함께한  
학주, 원찬 두 형님들과 잘 지도해주신 양홍석, 이광근 두 분 교수님들께 고마운 마음을 전한다.

## 인상깊은 논문

### Expressing and Verifying Probabilistic Assertions

Adrian Sampson, Pavel Panchekha, Todd Mytkowicz, Kathryn S. McKinley,  
Dan Grossman, Luis Ceze

최근들어 많은 주목을 받고 있는 확률적 프로그램 (Probabilistic Program)을 위한  
'assertion' 을 도입하자는 주제이다. 전통적인 프로그램에서 assertion은 그 지점에 도달하는  
모든 실행 경로에서 해당 식을 만족하는지를 검사한다. 하지만 이러한 방법으로는 확률적 프로그  
램의 실행을 제대로 기술하고 검사할 수 없다. 확률적 프로그램은 그 특성상 매번 다르게 실행되  
고, 다만 실행이 특정 분포를 따르기 때문이다.

이 논문에서는 'passertion' 이라고 부르는 확률적인 assertion으로 프로그램이 만족해야 하  
는 성질을 기술한다. passert 문은 프로그램이 실행되면서 만족해야하는 조건과 그 확률로 구  
성된다. 예를 들면,

```
x = random.uniform(0,1)  // 0 과 1 사이 수를 임의로 선택
w = 0.9
passert x < w, 0.90
```

위와 같은 프로그램에서 passert 문은  $x < w$  라는 조건이 90% 확률로 만족된다는 성질을 표현  
하고 있고, 이는 실제로 참이다.

이러한 확률적 성질을 검사하기 위한 아이디어는 간단하고 직관적이다. 프로그램을 실행 시킬  
때 확률적 값은 심볼로 취급하는 것이다. 그렇게되면 passert 의 조건에 해당하는 식은 심볼을  
포함한 식으로 구성될테다. 이 식은 곧바로 Bayesian Network 으로 변환할 수 있고, 기존에  
있는 각종 통계 기법들을 사용해서 식을 간략화 할 수 있다. 이 과정을 통해 간략해 진 식으로  
passert 문의 참/거짓을 판단하는 것이다. 하지만 식이 복잡한 경우에는 이를 판단하는 것이 쉽  
지 않은데, 그 때는 각 심볼이 표현하는 분포에서 직접 값을 샘플링 하여 통계학적 검증을 한다.

꼭 필요한 개념과 문제들을 처음 제시하고 아주 간단한 아이디어로 해결한 점이 마음에 드는 논  
문이다. 물론, 복잡한 경우에 잘 동작하지 않을지도 모르지만 이제 막 태생하는 분야에 꼭 필요한  
초석을 하나 놓은 점에서 높이 평가 받지 않았나 싶다.

---

## On Abstraction Refinement for Program Analyses in Datalog

Xin Zhang, Ravi Mangal, Radu Gregore, Mayur Naik, Hongseok Yang

프로그램 분석에서 원하는 성질을 검증하기 위해 가장 적절한 요약(abstraction)을 찾는 논문으로서, Mayur Naik 과 양홍석 교수님이 최근 몇 년간 계속 연구하는 주제이다. 프로그램 분석은 필연적으로 요약이 들어간다. 그런데 요약을 과하게 하면 정확도가 떨어져 원하는 성질을 검증하지 못하고 적게하면 시간이나 메모리 등 비용이 많이 들어서 문제다.

여기서는 Datalog 로 기술되는 특정 부류 분석에서 가장 적절한 요약을 찾는 방법을 이야기한다. 핵심 아이디어는 CEGAR (Counter-Example Guided Abstract Refinement) 를 이용하는 것이다. 즉, 가장 조악한 요약으로 먼저 프로그램을 분석한 후, 원하는 성질을 검증하지 못하면 그 원인을 살펴보고 좀 더 세밀한 요약을 선택해서 다시 분석한다. 이 때 해결해야 할 문제는 (유한하지만) 수많은 요약 가능성 중에서 어떤 것을 선택하느냐는 것이다. 여기서는 Datalog 분석의 특징을 잘 이용하여 이 문제를 MAXSAT (MAXimum SATisfiability) 문제로 치환해 풀었다. 이 방식으로 가장 적절한 요약을 찾아내거나 없다고 말해줄 수 있다.

적절한 요약을 찾기 위한 Mayur Naik 의 최근 몇년 행보는 대단하고 흥미롭다. PLDI 가 끝나고 SOAP 라는 위성 워크샵에서 Naik 의 초청 강연을 들었는데 4년에 걸친 여정을 한 번에 정리해 주었다. 최적 요약이 어느 정도 수준인지 잘 관찰한 다음, 특정 부류 분석에서 어떻게 효과적으로 그것을 찾는지 꾸준히 알아내고 있더라. 그리고 그 한 걸음 한 걸음이 많은 주목을 받으며 POPL 과 PLDI 에 줄기차게 채택되고 있다. 우리 연구실에서도 분석기의 정확도에 관심이 많은데, 우리가 대상으로 하는 분석에서 배울점이 있을지 살펴 보아야할 것 같다.

## Automatic Runtime Error Repair and Containment via Recovery Shepherd

Fan Long, Stelios Sidiropoulos-Douskos, Martin Rinard

MIT 의 Martin Rinard 교수님께서 꾸준히 연구하고 있는 ‘대충 계산 (approximate computing)’ 의 매우 실용적인 버전이다. ‘대충 계산’ 이란 프로그램 코드에 적혀있는 그대로 프로그램을 실행하는 것이 아니라 목적에 따라 대충 대충 계산하는 것을 말한다. 계산 결과가 부정확해도 상관없이 속도가 목적이면 프로그램의 명령어를 일부 생략하고 계산하여 결과를 낸다. 실시간 동영상 디코딩같이 완벽한 화질을 디코딩하는 것보다 사람이 답답해하지 않을 재생 속도가 중요할 때 유용하게 쓰일 수 있다. 그런 계산을 어떻게 특정 성질을 유지하면서 하는지 그 해결책에 관한 발표를 이전에 몇 번 들은적이 있다.

이번 논문은 계산하다가 오류가 나서 프로그램이 죽는 경우에도 그 부분만 무시하고 어떻게든 프로그램을 끝까지 실행시키는 데 초점을 맞춘다. 예를 들면, 웹브라우저가 웹페이지를 여는데 중간에 오류가 발생했다고 해서 죽지말고, 페이지 일부가 손상되더라도 페이지를 열게 만드는 것이다. 사용자에게는 그것이 다음 버전 웹브라우저를 기다리는 것 보다 낫다.

기술적인 부분은 아주 간단하다. 여기서는 0으로 나누는 오류와 Null 참조 오류를 대상으로 한다. 이 두 오류가 발생하면 특정 신호 (SIGFPE 와 SIGSEGV) 가 전달되는데 이를 가로막고 아무일 없었던냥 계산을 계속하도록 만든다. 이때 각 경우별로 미리 정해진 행동을 하는 것이다. 만약 0으로 나누는 오류가 발생하면 그 신호를 가로채고 나누기 결과값이 저장되어야 할 레지스터에 임의로 숫자 0을 채워 넣는 것이다. 이렇게 정해진 행동들은 실험적으로 구한 것이라 한다.

보통 프로그램 오류나 분석에 관한 논문은 개발자를 이롭게 하는 주제가 많은데 이는 사용자를 직접적으로 위한 기술이라는 것이 흥미롭다.

---

## FlashExtract: A Framework for Data Extraction by Examples

Vu Le, Sumit Gulwani

Sumit Gulwani 가 늘 연구하는 주제로서, 프로그램 합성을 이용해 사용자가 단순하고 많은 작업을 손쉽게 수행하도록 하는 논문이다. 이번에는 구조가 있는 문서 (텍스트, 웹페이지, 스프레드시트 등) 에서 전체 혹은 일부에 비슷한 작업을 사용자가 한 번에 하게끔 만드는 기술이다. 예를 들면, 이름, 학번, 나이, 주소가 적힌 사람 명단에서 모든이의 이름만 추출하고자 하는 경우, 한 명의 이름만 선택하면 나머지 모두의 이름이 자동으로 선택되어 추출할 수 있다. 프로그래밍을 잘하는 사람들에게는 간단한 스크립트로 짜면 일도 아니겠지만 그렇지 않은 일반 사용자에게는 굉장히 도움되는 기술일 것이다.

이 분야의 특성상 자세한 설명보다는 시연을 보여주면 모두가 어떤 기술인지 이해를 하기 때문에, 발표도 시연위주로 진행했다. 하지만 막상 시연을 하니 몇 가지 문제가 발생해서 원활히 진행되지 못했다. 그래서 미리 준비해둔 녹화 영상으로 대체했는데, 큰 발표에서는 예상치 못한 변수가 많고 시간도 부족하므로 시연보다는 동영상을 보여주는 것이 나을 수도 있겠다.

## Compiler Validation via Equivalence Modulo Inputs

Vu Le, Mehrdad Afshari, Zhendong Su

컴파일러의 버그 중에서도 특히 최적화 단계의 문제를 찾기 위한 기술인데 좀 특이하다. EMI (Equivalence Modulo Input) 라는 개념을 도입한 것이 핵심인데, 어떤 입력의 집합  $I$  에 대해서 결과가 똑같은 프로그램 집합을 지칭한다. 어떤 프로그램  $P$  에다  $I$  를 입력으로 주고 실행경로를 관찰한 후 수행되지 않은 명령어를 이리저리 바꾸면  $P$  와 EMI 관계를 갖는 프로그램을 생성할 수 있다. 이렇게 생성된 프로그램은  $I$  와 상관없는 부분을 바꾸었으므로 다시  $I$  를 넣고 돌려도  $P$  와 같은 결과를 내어야만 한다. 그런데도  $P$  와 다른 결과가 나왔다면 컴파일러에 버그가 있는 것이다.

이는 컴파일러 최적화에서 이용하는 정적분석의 특징을 잘 이용한 것이다. 정적 분석은 모든 입력 가능성을 고려하기 때문에 같은 EMI 클래스에 속한 프로그램이라도 분석결과는 확연히 다르게 나올 수 있다. 이 때문에 최적화 단계의 버그는 더 잘 드러내고, 비슷한 코드만 쳐다보는 개발자는 더 쉽게 버그를 고치거나 보고할 수 있다.

쉽게 떠오르지 않을것 같은 재미난 아이디어가 돋보인다. 다른 논문에 적힌 해결책에 비해서 덜 직관적이다. 그래서 그런지 논문에도 이 기술이 왜 잘 동작하는지 그 직관을 전달하려고 애쓴 부분이 많이 보인다. 어디서 어떻게 아이디어를 떠올렸는지 물어봤어야 했는데 그러지 못해 아쉽다.

## FlowDroid: precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for Android apps

Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Ochteau, Patrick McDaniel

안드로이드 개인정보 누출 분석을 위한 정적 분석기를 만든 논문이다. 무시무시한 제목에서 볼 수 있듯이 실제 쓸수 있는 탄탄한 구현체를 만들어서 공개했다. 개인정보 누출 분석이 유한 도메



인을 이용하는데다 의미함수가 간단한 특징을 이용하여 잘 알려진 IFDS 알고리즘을 활용하였다. 또한, 안드로이드의 특성에 따라 라이프 사이클을 다룰수 있게 하고, 효율적인 포인터 분석을 디자인 하는 등, 필요한 부분을 추가로 제시하였다.

크게 이론적으로 의미있는 점은 없지만 그럴듯한 분석기를 만들어 공개한 점에서 큰 인정을 받은 것 같다. 요즘 들어 여러 학회에서 AE(Artifact Evaluation) 를 시행하고 있는데 PLDI의 AE 위원회로부터도 선정된 논문이다. 이 프로젝트를 이끌고 있는 Eric Bodden 은 수년전부터 FlowDroid 에 사용된 모든 요소들을 만들어 공개하고 있었다. 자기 일에 자신감을 갖고 유용한 지식을 널리 제공하겠다는 의지가 없으면 힘든 일이다. 이전에, FlowDroid 의 앞단으로 사용되는 Soot 의 메일링 리스트에 가입한 적이 있었는데 많은 질문에 직접 답을 해주는 모습이 인상적이었다. 이런 움직임이 모여서 학계에서도 공개 소프트웨어 바람이 더 거세게 불고 있는 듯 하다. 좋은 일이다.

## 사람들

PLDI 를 비롯하여 몇 번 학회를 참가하고 대학원 생활을 하다보니 낯익은 얼굴이 많이 보였다. 이번 학회는 장소 문제 때문인지 교류를 위한 의도인지 모르겠지만 점심을 서서 먹게 했다. 고시촌 컵밥처럼 작은 밥그릇에 음식을 담아놓고 서서 먹는 방식이었다. 끝에가서는 좀 지쳤지만, 나름 맛도 있었고 덕분에 모르는 사람들과 몇마디 나눠볼 기회도 가졌다. 2009년에 우리 연구실을 방문했던 Will Klieber 를 오랜만에 만났는데 여전히 착하고 순수한 형이다. 올해 CMU 에서 박사학위를 받고 안드로이드 관련 연구를 계속 하고 있다 하였다. SOAP 에 안드로이드 개인 정보 누출 분석 논문을 발표했는데 재미있게 들었다. 홍콩과기대에 다니는 Xiao Xiao 는 2012 년 PLDI 에서 만나 우리의 스파스 분석에 많은 관심을 가던 친구였는데 이번에 저자로 다시 만나 반가웠다. 포인터 분석 정보를 효율적으로 데이터베이스에 저장하고 연산하는 기술을 발표하러 왔다고. 구글에 게시는 하정우 박사님도 지난 PLDI 에 이어 또 만나게 되었다. 구글에서 벌어



지는 일과 정적 분석의 필요성에 관해 좋은 말씀을 많이 해주셨다. 그 외에도 쉬는 시간에 만난 여러 사람들에게 우리 논문에 관해 간략히 소개하는 재미도 있었다. 매번 느끼는 것이지만 세상은 좁고, 학계는 더 좁아서 참 많은 인연이 생긴다.

## 소요유

한국에서 에딘버러까지 12시간 넘는 비행동안 읽을 책으로 “장자”를 선택했다. 그 첫번째 장이 “소요유” 인데 자유롭게 거닐며 노닌다는 뜻. 유럽의 생기넘치는 자연을 느끼며 멋진 거리를 거닐 생각이었고 즐거운 나날을 보냈다.

에딘버러는 한가운데에 있는 에딘버러 성을 중심으로 이루어진 도시이다. 성은 그리 크지 않았지만 절벽위에 우뚝 서 있어서 도심 어디에서나 훌륭한 볼거리가 된다. 시도때도 없이 비내리는 영국이지만 다행히 학회기간 전반부에는 날씨가 무척 맑아 성과 어우러진 하늘이 아주 근사했다.

에딘버러는 추모의 도시였다. 도시 곳곳에 동상이 서 있었다. 아담 스미스, 데이비드 흄 등 영국의 위인들이었다. 이와 더불어 공원의 각 벤치마다 기념하는 글이 하나씩 쓰여 있었다. 그런 벤치의 글은 위인을 추모하는 것은 아니고 에딘버러 사람들의 소소한 기억이었다. 돌아가신 아버지를 기린다거나, 심지어 무슨 생화학 연구소 몇주년 기념 이런 일상의 조각들. 벤치마다 부식정도가 다르고 줄도 맞춰져 있지 않은 것을 보아하니, 추측건대 추모나 기념을 하고 싶은 사람들이 공원에 벤치를 기증하면서 글을 적는 것이 아닌가 싶다.

음식이 맛없기로 소문난 대영제국인데, 맛이 없다기보다는 토속음식이 별로 없는 듯 했다. 길거리에는 이탈리아, 중국, 인도, 동남아 음식점들만 보였다. 스코틀랜드에는 ‘하기스 (Haggis)’ 소나 양의 내장을 다져서 위장에 넣고 삶은 토속음식이 있다. 조리법도 맛도 우리나라에서 먹는 찹쌀 순대와 매우 비슷하다.



## 마무리

학회 참가는 늘 연구에 새로운 원동력을 불어 넣는다. 좋은 논문을 쓸 수 있도록 잘 지도해주시고 학회 참가 지원도 해주신 이광근 교수님 덕분이다. 또한, 여러 가지 행정 절차를 도와주신 소프트웨어 무결점 연구센터 행정실 분들의 도움으로 편히 갔다올 수 있었다. 이제 다시 새로운 마음가짐으로 2014년 하반기에 뛰어들 때이다.

