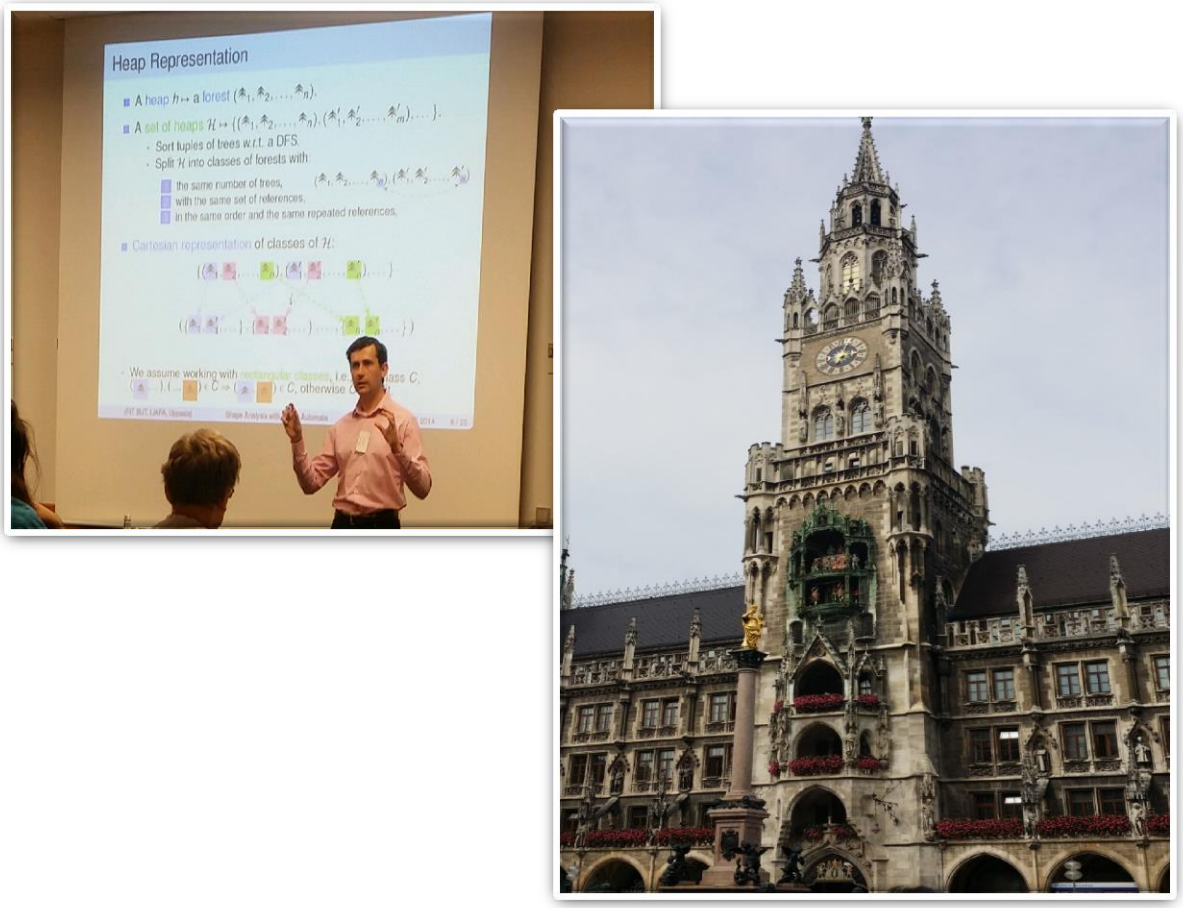
Static Analysis Symposium 2014



Munich, Germany

2014.09.09 – 2014.09.15

최재승

서울대학교 프로그래밍 연구실

들어가며

지난 9/9 ~ 9/13 에 독일 뮌헨에서 열린 Static Analysis Symposium 2014 에 다녀왔다. Static Analysis Symposium (SAS)은 프로그램 정적 분석과 관련된 주제로 다양한 논문이 발표된다. 연구실 선배이신 이우석 형은 "Static Analysis Progress Bar" 논문이 채택되어 발표를 위해 함께 참석하셨으며 이에 동행하여 다녀오게 되었다. 처음으로 참석해 보는 학회였는데, 많은 것을 경험할 수 있는 뜻깊은 기회였다.

TAPAS 발표

SAS 학회가 진행되기 전 하루 동안은, 위성 워크샵인 Tools for Automatic Program Analysis (TAPAS)가 진행되었으며, 위성 워크샵이었음에도 흥미로운 발표가 많았다.

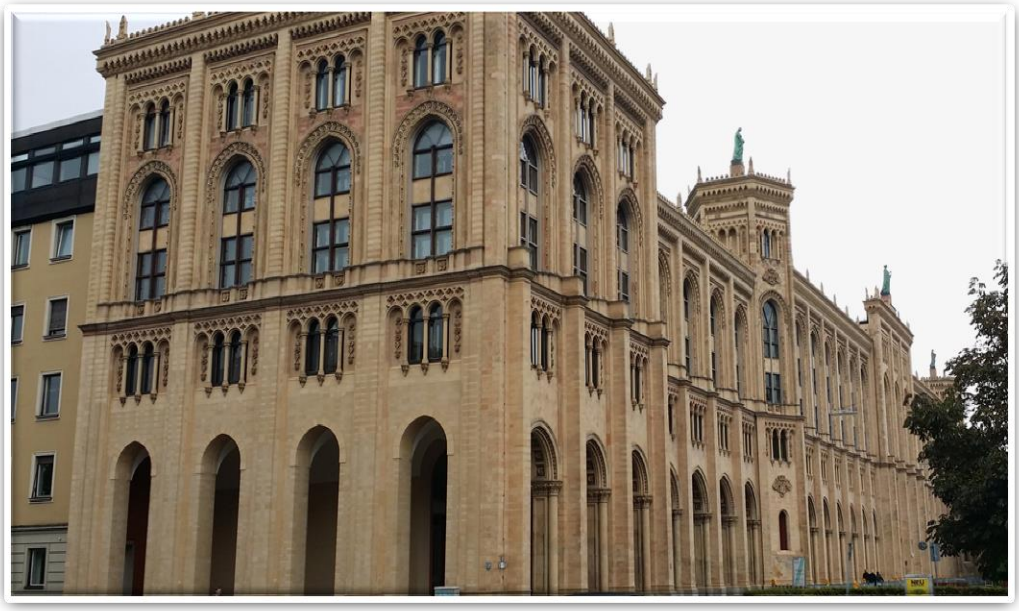
Machine-code analysis and transformation at GrammaTech

B.Reps, Gogul과 함께 연구 및 개발중인 바이너리 분석기인 CodeSurfer 및 CodeSonar에 대한 발표였다. CodeSurfer는 요약 해석에 기반한 분석기로서, Strided interval 공간을 사용하여 바이너리 프로그램의 메모리 상태를 분석한다. 그리고 이 분석기를 보안 취약점 탐지를 위해 튜닝한 것이 CodeSonar이다. 나도 같은 주제를 연구하고 있기 때문에 더욱 관심을 갖고 세밀한 부분까지 주의를 기울이며 발표를 들었다. 이전에 Analyzing Memory Access in x86 Executable 논문으로 읽었던 내용이 잘 정리되는 효과도 있었으며, 논문에는 언급되지 않은 내용이나, 그 이후로 어떤 방향으로 후속 연구가 이뤄지고 있는지를 들어볼 수 있어 의미있는 경험이었다. 발표를 들으면서, 내 연구가 어떤 차별점과 의의가 있는지 꼼꼼하게 따지며 되짚어 보는 기회가 되었다,

Declarative Static Program Analysis

프로그램의 포인터 분석을 위한 general한 프레임워크인 "Doop"에 대한 발표였다. Pointer analysis에 대해서는 그 간단한 개념만 알고 있었으며, 구체적인 발표를

들어보는 것은 이번이 처음이었는데, 익숙하지 않은 나 같은 청중에게도 핵심적인 개념을 잘 전달해내는 발표 능력에 깊은 인상을 받았다. Doop은 포인터 분석을 위한 여러 알고리즘을 표현하는 일반적인(general) 프레임워크이며, 다음과 같은 특징이 있다. 우선 알고리즘을 표현하는 방식 면에서는 declarative한 스타일로 간결하게 표현할 수 있으며, 잘 parameterized되어 있어 flow sensitivity, context sensitivity를 굉장히 쉽게 표현할 수 있다. 또한, 이렇게 작성한 algorithm은 내부적으로 database engine에 기반하여 구현되는데 분석 속도가 1.2배 정도 빨라진다고 한다. 포인터와 location 사이의 points-to 관계를 database 이론으로 해석하고 구현한 것이 신선했고, 자신이 개발한 프레임워크의 유용성을 남들에게 설득하기 위한 논리는 어떻게 전개해야 하는지에 주의를 기울여 들었다.



Invited Talks (in SAS)

3일동안 진행된 SAS에는 매일 한 명씩의 연사가 자신의 연구 내용을 발표하는 Invited Talk가 준비되어 있었다.

Dynamic Program Verification

첫째 날은 Microsoft의 SAGE 개발을 주도한 Patrice Godefroid가 프로그램의 동적

인 검증 (dynamic verification)을 주제로 발표하였다. Concolic tester인 SAGE에 대해서, 그리고 나아가 이를 동적인 검증으로 발전시키기 위한 노력에 대해 설명하였는데, 수동으로 찾아낸 프로그램의 성질과 concolic testing을 결합하여, ANI 포맷 파일 파서의 메모리 안정성(safety)을 검증한 case study를 소개하였다. 이런 verification이 순수하게 동적으로 이뤄질 수 있는지에 대해서는 청중과 의견 차이를 보였는데, 특히 Cousot의 날카로운 비판이 인상깊었다.

Fully Automated Shape Analysis Based on Forest Automata with Data Constraints

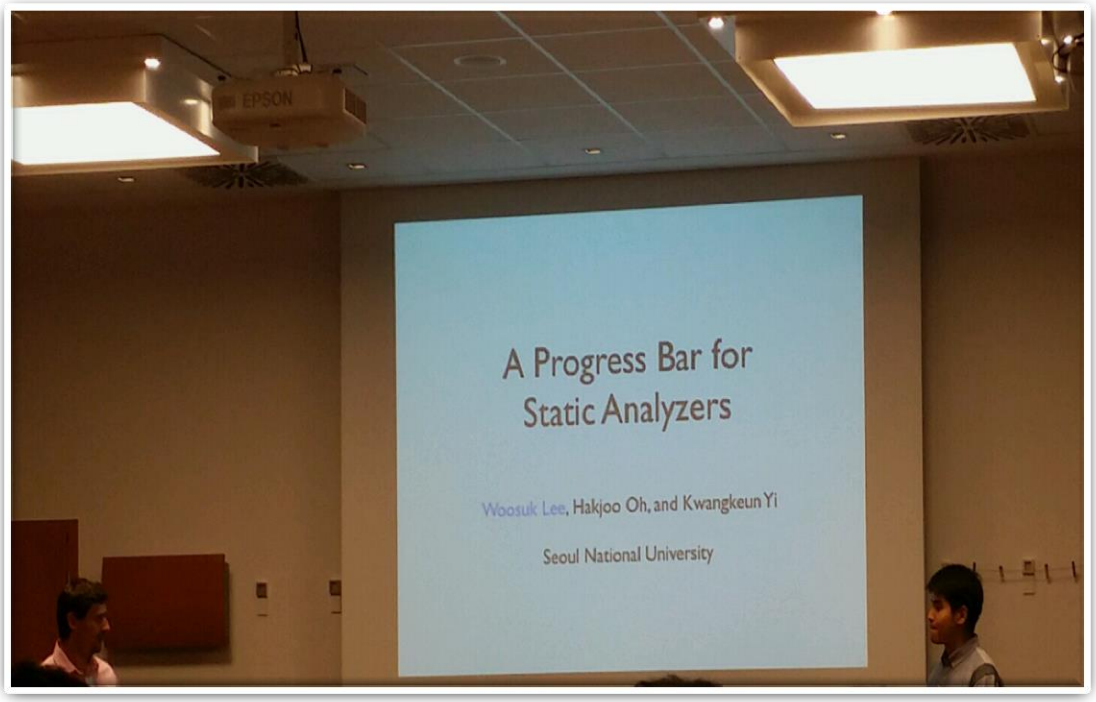
마지막 날에는 Tomas Vojnar가 오토마타 이론에 기반한 shape analysis에 대한 발표를 진행하였다. 최근 몇 년 동안 보안 분야에서 mangling reference 문제가 보안 분야에서 심각한 이슈로 떠오르고 있었기에 많은 관심을 가지고 있었으나, 힙 메모리와 관련된 버그를 잡아낼 수 있는 shape analysis에 대해서는 정작 거의 모르고 있었다. 이 발표의 내용은 이해하기 벅찬 부분이 굉장히 많았지만, shape analysis를 위한 여러 기법 중에 오토마타를 사용한 기법에 대한 좋은 맛보기가 되었던 것 같다. 공부하려면 많은 노력이 필요한 주제이겠지만, 틈틈이 시간을 들여 기초적인 부분부터 공부해 보고 싶다는 흥미를 느꼈다.

SAS 발표

A Progress Bar for Static Analysis

정적 분석이 얼마나 진행되었는지를 유추하는 방법을 최초로 제안하는 논문으로써, 연구실의 이우석 선배가 발표한 논문이다. 무엇보다도 신선하고 유용한 문제 의식 때문에 많은 청중들의 관심을 받았으며 질문 시간에도 많은 질문을 받았다. 이번 학회에서는 굉장히 엄격하고 비판 및 문제제기를 받은 발표가 여럿 있었는데, 다행이 이 발표는 좋은 반응을 받았으며, 질문 시간에도 그 다양한 응용 방법에 대한 제안을 많이 들을 수 있었다.

프로그램 분석에서 진행도를 추측하는 것은 쉽지 않은 일인데, 프로그램의 값이 lattice를 타고 올라가는 속도가 일정하지 않으며, 얼마나 높이까지 올라가야 fixpoint에 도달하는지를 미리 알기 힘들기 때문이다. fixpoint의 lattice에서의 높이를 유추하기 위한 전분석을 도입하는 것을 시작으로 하여, 문제를 하나씩 차근차근 해결해 나갔다. 발표 슬라이드는 각 장마다 말하려는 바가 쉽고 뚜렷하게 나타나 있어 전달력이 매우 좋았다.



Sparse Dataflow Analysis with Pointers and Reachability

javascript 언어에서 sparse 분석을 하기 위해 on the fly로 def-use chain을 그리며 분석하는 기술에 대한 발표였다. 이 발표에 관심을 가졌던 이유는, 우리 연구실에서 Sparse Analysis를 설계하기 위한 일반적인 틀을 연구하여 발표한 바 있는데, 이와는 다른 방법으로 문제에 접근하였기 때문이다. 왜 우리 연구실의 방법을 적용하지 않았는지가 논문에 너무 간단하게만 언급되어 있어서, 학회 틈틈이 계속 머리를 굴려야 했다. 연구실 선배님들이 정립한 방법은, 각 변수가 어떤 location을 가리키는 지에 대한 정보를 흐름 둔감(flow insensitive)한 예비분석의 결과로부터 가져온다 (PLDI 논문의 Section 3.2). 그런데, javascript 언어로 작성된 프로그

램은, 문맥에 따라 이 location 정보가 극적으로 바뀌는 특성이 있다고 한다. 이로 인해, context sensitivity가 없는 예비 분석으로 data dependency를 구하면 분석이 별로 sparse해지지 않는 현상이 발생한다. 이를 해결하기 위해 발표자는 분석하는 도중에 def-use chain을 수정해 나가면서 sparse한 분석을 진행하는 방법을 제안하였다. 이 방법을 적용했을 때, 가장 정확한 최적의 DU chain 주어진 sparse 분석과 비교해서도 2배 이상 느리지 않았으며, 이는 on the fly로 DU chain을 그리는 것의 오버헤드가 크지 않았음을 의미한다. 다만, 어떻게 알고리즘이 작동하는지와 그에 대한 직관적인 설명은 논문에 자세히 쓰여 있었으나, 그 방법이 올바르다는 것을 체계적으로 증명한 내용이 없었던 점이 아쉬웠다.

Region-based Selective Flow-Sensitive Pointer Analysis

이 발표는 포인터 분석을 위해 부분적으로 흐름에 민감한 (selective flow sensitive) 분석을 하는 방법을 제시하였다. 프로그램 포인트를 여러 지역(region)으로 나누고, 각 지역 안에서는 flow insensitive한 분석을 하고, 지역 사이에서는 flow sensitive한 분석을 한다. 이 때 분석의 속도 및 정확성을 결정하는 것은 지역을 나누는 방식이 되며, 여기에 어떤 알고리즘을 쓰느냐에 따라 속도와 정확성 사이의 교환(trade-off)이 일어난다. 이 논문에서는, 모든 Load (포인터 참조) 연산의 flow sensitive한 분석 결과의 정확성을 유지하도록 지역을 나누는 알고리즘을 소개하였다. 이를 위해 예비 분석을 통해 얻은 points-to 정보를 사용하는데, 구체적인 알고리즘은 아직 틈틈이 들여다보는 중이다. 이렇게 지역을 나누는 방식을 사용할 경우, 정확도는 flow sensitive에 가까워지지만, 속도 향상은 평균적으로 2.11배에 그친다. 우리 연구실의 기술인 효과 예비 분석 (impact pre-analysis)를 사용하는 방식은 precision을 희생해서 많은 속도 향상을 얻는데, 이와 대조되는 결과이다. 같은 문제에 대해 다른 접근 방식으로 문제를 해결하는 두 사례를 살펴보는 기회가 된 것 같아 흥미롭다.

SawjaCard: a Static Analysis Tool for Certifying Java Card Applications

NFC 등의 카드 칩을 프로그래밍하는데 사용되는 Java 기반 언어인 Java Card로 작성된 프로그램을 분석하는 툴에 대한 발표였다. 분석하려는 성질에 맞추어, 기존에 알려진 여러 요약 공간들을 선택 및 조합하거나 필요에 따라 새로운 공간을 정의하여 사용하였다. 이 발표에 관심을 가졌던 이유는, 내가 연구하고 있는 주제 또한 기존에 잘 알려진 요약 해석의 이론을 binary executable이라는 언어에 대해 적용하는 주제이기 때문이다. 이론적으로 새로운 것을 발견해내는 것이 아닌 응용적인 연구를 할 때, 그 의미를 인정받기 위해서는 어떤 것이 갖춰져야 하는지에 유의하며 발표를 들었는데, 이 부분은 앞으로도 지속적으로 고민해 봐야 할 것 같다.

사람들

학회에서 좀 더 많은 사람들과 안면을 트고 이야기를 나누었으면 좋았을 텐데 하는 아쉬움이 남는다. 처음 참석하는 학회라 긴장했던 점도 있고, 영어로 대화하는 속도가 빨라 내용을 제대로 알아듣지 못했던 이유도 있는 것 같다. 연구 내용에 대한 대화가 오갈 때에는, 배경지식이 부족하여 이해하지 못하는 경우도 있어 아쉬웠다.

그럼에도, 대부분의 사람들이 친절하고 우호적이어서 대화에 참여할 의욕을 낼 수 있었는데, TAPAS에서 바이너리 분석기에 대해 발표한 Alexey Loginov와 같이 점심식사를 하며 문제에 어떻게 접근하고 있는지 간략히나마 들어볼 수 있어 좋은 기회가 되었다.

이번 학회에 참석한 한국인은 우리 연구실을 제외하고는 아무도 없었는데 이 부분은 조금 아쉬웠다. 중국 국방과학기술대학교의 Liqian Chen을 이우석 선배가 소개해 주셨는데, 관계분석을 위한 다양한 연구 공간을 연구하는 분이셨다.

생활

뮌헨의 날씨는 서울에 비해 제법 쌀쌀했으며, 특히 비가 올 때는 굉장히 서늘하여 옷을 세 겹씩 입고 다녔다. 일반적으로 알려진 바로는 서울과 날씨가 비슷하다고 하였는데, 혹시 모른다는 생각에 미리 뮌헨의 일기예보를 살펴보았기에 따뜻한 옷을 준비할 수 있었다. 앞으로도 미리 출장 기간 동안의 날씨를 알아보고 가는 것이 안전할 것 같다.

학회장 그리고 뮌헨에서 둘러볼 만한 대부분의 곳이 도시 중앙을 가로지르는 지하철로부터 가까이 자리잡고 있어 수월하게 이동할 수 있었으며, 지하철은 공항까지 이어져 있다. 독일의 지하철은 역 안의 펀칭 기계에 스스로 표를 찍고 들어가는 방식이며, 굉장히 허술하고 사용자의 양심에 의존하는 시스템이다. 다만 사복을 입은 검표원이 불심 검문을 하여, 무임승차가 적발될 경우 큰 액수의 벌금을 물린다고 한다.

뮌헨의 둘러볼 만한 곳은 고딕 양식으로 정교하게 지어진 시청 건물 (지하철역 마리엔 광장 근처) 근처에 모여 있는 편이다. 나란히 높이 솟아 있는 돔 지붕의 탑이 인상적인 프라우엔 교회, 길거리 악사들과 자주 마주칠 수 있었던 노이하우어 거리, 유명한 비어홀 등이 자리잡고 있다.

식사는 학회장에서 제공해 주는 식사에 참여하거나, 독일의 유명한 음식들을 찾아다니며 해결하였다. 독일의 유명한 음식인 소시지나 학센(독일식 족발)과 같

은 음식들은 다행히 무난하게 즐길 수 있었으나, 마지막 날 정도 되자 기름진 고기 위주의 식사가 다소 부담스럽기도 했다.

마치며

Ropas 연구실에서 공부하기 시작한 뒤로 처음으로 다녀온 학회였는데, 굉장히 뜻깊은 한 경험이 되었던 것 같다. 자신이 연구한 내용이 학회라는 장소에서 엄밀하고 냉정한 잣대로 평가되고 피드백을 갖는 과정이 굉장히 인상깊으면서도, 도전 의식을 불러일으켰다. 또한 요약 해석에만 시선을 고정할 것이 아니라, 모델 체킹이나 분리 논리(separation logic)과 같은 기술들도 폭넓게 접하고 공부하여, 다음에 이런 자리에 참석했을 때는 더 많은 지식과 아이디어를 배워올 수 있도록 해야겠다고 생각했다.

학회에 참석해서 유익한 경험을 할 수 있도록 지원해 주신 이광근 교수님, 소프트웨어 무결점 연구센터 및 행정팀에 깊이 감사드리며, 좋은 기회를 받은 것이 헛되지 않도록 학회에서 받은 자극을 원동력 삼아 열심히 연구하고 싶다.