Programming Languages and Technical Disruption

Ben Zorn
Principal Researcher and Research Manager
Research in Software Engineering (RiSE) Group
Microsoft Research, Redmond
This talk is about...

Well-being

Why Zika Is This Year’s Scary Virus
It is “spreading explosively” in the Americas and may be the next public health emergency.

Stability

Microsoft's Excel Might Be The Most Dangerous Software On The Planet

Tim Worstall, CONTRIBUTOR
I have opinions about economics, finance and public policy. FULL Opinions expressed by Forbes Contributors are their own.

No, really, it’s possible that Microsoft’s Excel is the most dangerous software on the planet. Yes, more dangerous than rogue code running a nuclear power plant, that was deliberately sent off to sabotage Iran’s nuclear program, worse, even, than whatever rent in the fabric of space-time led to the invention of Lojacks. Really, that serious.


Trust

Volkswagen: The scandal explained
By Russell Hotten
Business reporter, BBC News
© 10 December 2015

Zorn, PLDI 2016
RiSE – Research in Software Engineering

http://research.microsoft.com/en-us/groups/rise/
What’s New in Computer Science?

Deep Learning

Where are the Programming Languages?
Presenting DeepPLDI:
The First PLDI paper written by an RNN

- Data: 7 years of PLDI 2007-2015, 304 files, 25Mb of text
- HW: Intel i7, Nvidia GeForce GTX Titan GPU
- RNN: 3-layer LSTM, 512 cells/layer, 40k training batches
- Thanks to Bill Zorn, U. Washington
- Also see “The Unreasonable Effectiveness of Recurrent Neural Networks” http://karpathy.github.io/2015/05/21/rnn-effectiveness/
Presenting **DeepPLDI**: The First PLDI paper written by an RNN

Excerpt: “This work are not a solution to separation logic since the system requires only symbolic execution to erroridy details of dynamically nested query hardware. **We present a convenient alias analysis that attains a number of queries in a typedef style**, the design of the stand-representation is the case that symmetric constructs can be applied and how to matter. For example, inductive properties to support table loads and analysis, a keyword assumption to the signature for $\Gamma F$ across a dynamically computed list. Our language’s defects by computing these points are synthesized ing a synthesis of the sizes and mutations fence in Section 3.2.1.”

- DeepPLDI, 2016
Languages are Central to Disruption

• VisiCalc, 1979
• Visual Basic, 1991
• Java, 1991
• JavaScript, 1996
And Computing is Central to Everything...
The Rapidly Expanding World of Computing

- Scientific Discovery
- Transportation
- Neural Engineering
- Elder Care
- Medicine and Global Health
- Energy and Sustainability
- Security and Privacy
- Technology for Development
- Interacting with the Physical World

CORE
- CSE
- Natural language processing
- Sensors
- Mobile
- HCl
- Machine learning
- Cloud computing
- Big data

Graphic: Lazowska
Three Examples

Public Health
Financial Stability
Cybersecurity
Public Health


Zorn, PLDI 2016
Detour: Is Code Really Dead?

In the sciences the authority of thousands of opinions is not worth as much as one tiny spark of reason in an individual man.

Galileo Galilei (trans. 1957)

Third letter on sunspots (December 1612)
Honoring Automated Reasoning at Scale

Pioneering achievements
Changing math & science
Awards / recognition

Formal Proof—The Four-Color Theorem
Georges Gonthier

Automation of Mathematical Induction as part of the History of Logic
J Strother Moore
Dept. Computer Sci., Gates Dell C., 2317 Speedway,
The University of Texas at Austin, Austin, TX 78701
moore@cs.utexas.edu

Claus-Peter Wirth
FB AI, Hochschule Harz, D-38855 Wernigerode, Germany
wirth@logic.at
SEKI Report SR-2013-02

Nqthm

Z3
Z3 wins 2015 ACM SIGPLAN Award

Awards / recognition

ACM Software System Award

Zorn, PLDI 2016
Example: the Z3 SMT Solver
Leonardo de Moura and Nikolaj Bjørner, MSR

Many important problems can be expressed as a system of constraints in some logic

\[ x^2 + y^2 < 1 \text{ and } xy > 0.1 \]
\[ x^2 + y^2 < 1 \text{ and } xy > 1 \]

Is execution path \( P \) feasible?

Is assertion \( X \) violated?

Solution/Model

\[ \text{sat, } x = \frac{1}{8}, \ y = \frac{7}{8} \]

\[ \text{unsat, Proof} \]

Path Exploration

Program Verification

Z3 is a platform that many build on:
- 30,000 downloads
- 2800+ citations

De Moura, Leonardo, and Nikolaj Bjørner. "Z3: An efficient SMT solver. TACAS 2008"
SMT Solvers are already superhuman

Part of a 4-megabyte Z3 input generated in the proof of a verified TLS implementation (courtesy of Nikhil Swamy)
What does all this have to do with Public Health?
Public Health meets Formal Methods

What does this... Zika outbreak fuelled by mosquito control failure, says WHO boss

...have to do with this?
It starts with a question...

What if you could use mosquitos as sensors to detect the presence of infectious disease around the globe?
Building a Better Mosquito Trap

Catches individual insects
Uses wing frequency to identify species
Lightweight (5lb)
12-14 hour endurance
Reusable

C02-baited CDC UV trap, circa 2015
Premonition trap, 2016

Images courtesy of Ethan Jackson
Internet of (Field Biology) Things: Premonition

Analysis identifies Infectious diseases

DNA samples sent to cloud

Mosquito trap located in likely spots

Drone identifies placement sites

Repeat
Microsoft Research
Safe Cyber-Physical Systems Expedition

High-level Planning
Correct Control
Robust Sensing
Secure OS

Safe despite limited power, external disturbances, sensor noise and complex missions

Slide courtesy of Jeannette Wing
Automated Reasoning at Scale is Transformative

Enabling rapid advances in sciences based on:
  - Exponential improvements in sensing + devices
  - Cloud computing
  - Statistical and logical reasoning at scale

The future looks bright!
Financial Stability

Increasingly, society depends on correct and efficient analysis of data

It all started with spreadsheets...
Aside: Code, Data, and Noise

• Programming languages people like to think about code

• Consider Niklaus Wirth:

\[ \text{Algorithms + Data Structures = Programs} \]

Wrong!

\[ \text{Programs = Algorithms + Data Structures + Data} \]
Garbage in, garbage out is not the complete story

More reasoning required
Code and Data Always Coexist...

• Question (too easy): consider this regular expression

```regular_expression
([a-z0-9][-a-z0-9_\+\.]*)[a-z0-9]@([a-z0-9][-a-z0-9\.]*)[a-z0-9]\.
(arpa|root|aero|biz|cat|com|coop|edu|gov|info|int|jobs|mil|mobi|museum|name|net|org|pro|tel|travel|ac|...)
|([0-9]{1,3}\.{3}[0-9]{1,3})
```

• What does it do?
• Here’s a hint: benzorn@acm.org
Programming by example combines (implicit) code and data directly

In the PowerShell ConvertFrom-String commandlet:

```bash
{username*::benzorn}@{domain*::acm.org}
```

generates a program that generalizes to extract the user and domain name from every row of an input file

See: “Spreadsheet Data Manipulation using Examples”, CACM 2012, Research Highlights Paper, Sumit Gulwani, William Harris, Rishabh Singh and
https://blogs.msdn.microsoft.com/powershell/2014/10/31/convertfrom-string-example-basedtext-parsing/
Spreadsheets combine data and code in novel and innovative ways


<table>
<thead>
<tr>
<th></th>
<th>Annex 5.1</th>
<th>Final table: Historical trends in forest resources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1</td>
<td>Annex 5.1</td>
<td>Final table: Historical trends in forest resources</td>
</tr>
<tr>
<td>2</td>
<td>Forest area (1000 ha)</td>
<td>value</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Spreadsheets Have Bugs, Often About Money

Is This Spreadsheet a Tax Evader?
How H. M. Customs & Excise Test Spreadsheet Applications

Raymond J Butler, CISA
H. M. Customs and Excise, Ralli Quays, 3 Stanley Street, Salford, M60 9LA UK
rbutler.c&e.cau@gtnet.gov.uk / ray.butler@hmce.gov.uk
©Crown Copyright reserved, published by permission of the Commissioners of H M Customs & Excise

Findings:
- 7 files, 21 worksheets examined
- Average 713 formulas/file
- Total taxes: 12M pounds
- Total error: 1.37M pounds = 11% of total
Sometimes the Errors are Significant

JPMorgan Pays $920 Million to Settle London Whale Probes

2012 trading activities of a single individual at JP Morgan resulted in a total loss estimated at $6.2 billion

The actual problem in the Whale’s spreadsheet: “After subtracting the old rate from the new rate, the spreadsheet divided by their sum instead of their average, as the modeler had intended. This error likely had the effect of muting volatility by a factor of two and of lowering the VaR . . .”
Important Policies Are Based on Incorrect Computation

Researchers Finally Replicated Reinhart-Rogoff, and There Are Serious Problems.

By Mike Konczal | 04.16.13

Used by economists to justify austerity policy

Based on faulty calculations in the spreadsheet

http://rooseveltinstitute.org/researchers-finally-replicated-reinhart-rogoff-and-there-are-serious-problems/
Denmark, Canada, Belgium, etc probably feel that they should be included

<table>
<thead>
<tr>
<th>Country</th>
<th>Coverage</th>
<th>Debt/GDP</th>
<th>Real GDP growth</th>
<th>Debt/GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>30 or less</td>
<td>30 to 60</td>
<td>60 to 90</td>
</tr>
<tr>
<td>UK</td>
<td>1830-2009</td>
<td>3</td>
<td>68</td>
<td>27</td>
</tr>
<tr>
<td>Sweden</td>
<td>1880-2009</td>
<td>79</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>Spain</td>
<td>1880-2009</td>
<td>26</td>
<td>53</td>
<td>47</td>
</tr>
<tr>
<td>Portugal</td>
<td>1880-2009</td>
<td>9</td>
<td>32</td>
<td>39</td>
</tr>
<tr>
<td>Norway</td>
<td>1880-2009</td>
<td>13</td>
<td>26</td>
<td>47</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1912-2009</td>
<td>9</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1880-2009</td>
<td>17</td>
<td>50</td>
<td>32</td>
</tr>
<tr>
<td>Japan</td>
<td>1885-2009</td>
<td>47</td>
<td>42</td>
<td>11</td>
</tr>
<tr>
<td>Italy</td>
<td>1880-2009</td>
<td>26</td>
<td>12</td>
<td>39</td>
</tr>
<tr>
<td>Ireland</td>
<td>1880-2009</td>
<td>3</td>
<td>26</td>
<td>47</td>
</tr>
<tr>
<td>Denmark</td>
<td>1880-2009</td>
<td>9</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Canada</td>
<td>1925-2009</td>
<td>3</td>
<td>52</td>
<td>23</td>
</tr>
<tr>
<td>Belgium</td>
<td>1835-2009</td>
<td>37</td>
<td>60</td>
<td>32</td>
</tr>
<tr>
<td>Austria</td>
<td>1880-2009</td>
<td>43</td>
<td>32</td>
<td>35</td>
</tr>
<tr>
<td>Australia</td>
<td>1902-2009</td>
<td>38</td>
<td>33</td>
<td>23</td>
</tr>
</tbody>
</table>

Minimum

Maximum
Where are Formulas Expected?

Implicit dependence – formula expected
Spreadsheet Tools Still Underinvested in PL

Correctness advances
- Spreadsheet smells [Hermans et al., ICSM 2012]
- Data debugging / CheckCell [Barowy et al., OOPSLA 2014]
- Cell clustering / Custodes [Cheung et al., ICSE 2016]

Productivity advances
- Excel 2013 Flash Fill [Gulwani POPL 2011]
- Table extraction / FlashRelate [Barowy et al., PLDI 2015]
- BlinkFill [Singh VLDB 2016]
Correctness Beyond Spreadsheets

How many financial decisions are now based on technology that we don’t understand and is likely to have bugs?
Reasoning about Code + Data

Combining code with data is powerful:
  Spreadsheets
  HTML + JavaScript + CSS

Reasoning about data is as important as code
Tools for such systems are relatively immature
Much is at stake
Trust and Security
Recall: Rapidly Expanding World of Computing

- Medicine and Global Health
- Energy and Sustainability
- Security and Privacy
- Technology for Development
- Interacting with the Physical World

- CORE CSE
  - natural language processing
  - sensors
  - big data
  - cloud computing
  - machine learning
  - HCI
  - mobile

- Education
- Scientific Discovery
- Transportation
- Neural Engineering
- Elder Care
- Accessibility

Graphic: Lazowska
Observation:
Every company is a software company
Every object is a computer

Yes, this is a computer too

Microsoft

Zorn, PLDI 2016
Implications of CS + X

How much does your life depend on correct software?
What’s the trend?
Can governments understanding or regulate this trend?
When do you start worrying?
One Trend: Ransomware

Businesses and individuals increasingly targeted by ransomware

Dramatic increases in 2016

Smart Everything Increases the Attack Surface

Findings:
- Trivial credentials
- No encryption
- Easy to crash
- No longer supported

I bought some awful light bulbs so you don’t have to

Feb. 24th, 2016 04:37 PM

MJG59

I maintain an application for bridging various non-Hue lighting enough like a Hue that an Amazon Echo will still control them was colour support, so I picked up some cheap bulbs and a bri iRainbow001, and it’s terrible.

https://mjg59.dreamwidth.org/40397.html

Zorn, PLDI 2016
Did I Mention Companies Cheat?

You can hide a lot in 100M lines of code

Many Cars Have a Hundred Million Lines of Code

Who gets to write it?

December 3, 2012
Sounds really hard...
Required: Adversarial Thinking for All Engineers

Sokwoo Rhee (AD CPS, NIST): “How would you have prevented StuxNet?”

• Computer scientist: “Reduce attack surface, limit access, …”

• Mechanical eng.: “Install a limit switch on the centrifuges.”
Trust, but Verify

UL Launches Cybersecurity Assurance Program

New UL 2900 Series of Standards Offer Testable Cybersecurity Criteria for Network-Connectable Products & Systems

http://www.ul.com/cybersecurity

Emphasis on up-to-date patches, testing methods, code analysis, crypto

Do we really know how to provide “cybersecurity assurance”? How assured are you?
Can we really verify?
Two Expeditions: DeepSpec and Everest

Scalable reasoning meets software verification at scale

http://deepspec.org/

$10M NSF Expedition in Computing
Awarded 2016
Everest Expedition: Microsoft Research

Goal: drop-in verified HTTPS replacement

Challenges:
- scalability of verification
- performance
- usable tool chain

---

Slide courtesy of Cedric Fournet
Is Software the Hero or the Villain?
Disruptions are Happening

- **Smart objects** will replace dumb objects
- The software embedded in these objects will be written in the next 5 years, but will have implications for the next 50 years
- Our lives will depend on these objects...
- We need languages, tools, and processes to make these objects safe
The Cathedral and the Skyscraper

Heroic effort, amazing engineering, one of a kind...

Stronger materials, reusable components, mathematical analysis...
Building the Skyscraper: Key Elements

Reasoning at scale

Understanding code and data

Assuming cyber-physical systems

Requiring verified components
Thank you!

Acknowledgements
Daniel Barowy, Emery Berger, Nikolaj Bjorner, Shabnam Erfani, Cedric Fournet, Sumit Gulwani, Ethan Jackson, Ed Lazowska, Daan Leijen, Jonathan Protzenko, Sokwoo Rhee, Rishabh Singh, Nikhil Swamy, Emma Tosch, Moshe Vardi, Jeannette Wing, Bill Zorn, Sue Zorn, members of the CCC, MSR Safe CPS Expedition, MSR Everest Expedition

Follow me: @benzorn  https://twitter.com/benzorn

DeepPLDI: In order to guarantee termination of pairs defined via conversion and verification constraint from the valid-register files to termination and refinement types...
The Existential Threat of AI

Stephen Hawking, Elon Musk, and Bill Gates Warn About Artificial Intelligence

The Observer, 8/19/2015 http://bit.ly/1Pcjf1s

Rapid rise of AI in robotics, vision, etc. raises concerns
The Existential Threat of AI: Bad Software

From: http://blog.castsoftware.com/when-good-software-goes-bad/
The Existential Threat of AI - Bad Software

• AI is always part of a complete SW system
• Real risk is not knowing what SW can or will do...but:

• Assuming an adversary is first step (bad actors)
• Testing is the traditional approach
• Testing might fail when input/output behaviors are sufficiently complex – **this is a PL challenge**
Thank you!

Acknowledgements

Daniel Barowy, Emery Berger, Nikolaj Bjorner, Shabnam Erfani, Cedric Fournet, Sumit Gulwani, Ethan Jackson, Ed Lazowska, Daan Leijen, Jonathan Protzenko, Sokwoo Rhee, Rishabh Singh, Nikhil Swamy, Emma Tosch, Moshe Vardi, Jeannette Wing, Bill Zorn, Sue Zorn, members of the CCC, MSR Safe CPS Expedition, MSR Everest Expedition

Follow me: @benzorn https://twitter.com/benzorn

DeepPLDI: In order to guarantee termination of pairs defined via conversion and verification constraint from the valid-register files to termination and refinement types...