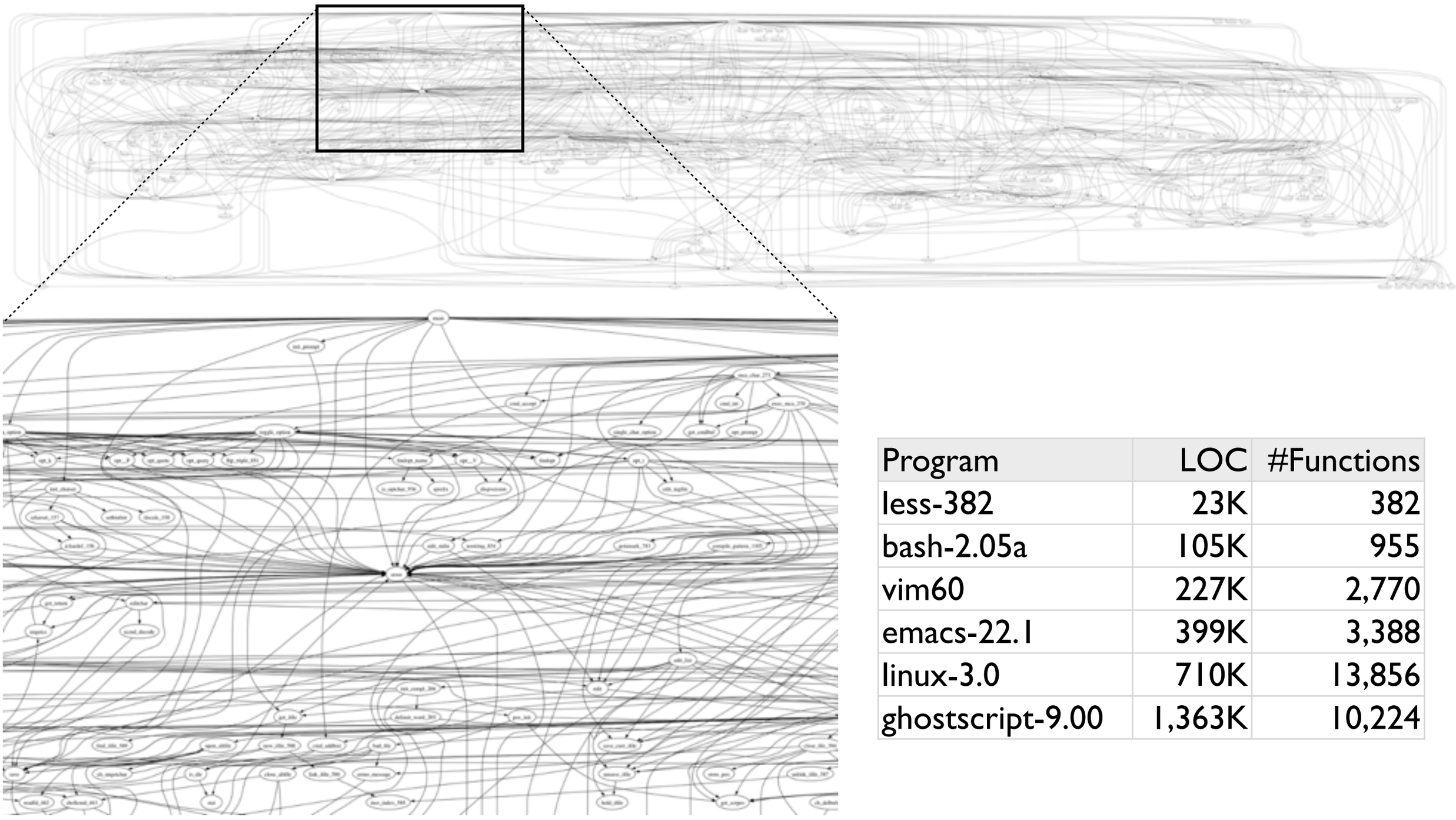
Global Static Analysis of Million Lines of Code

Hakjoo Oh
Seoul National University

Student Lightning Talk at POPL'12

Real-world Programs

less-382 (23,822 LOC)



Program	LOC	#Functions
less-382	23K	382
bash-2.05a	105K	955
vim60	227K	2,770
emacs-22.1	399K	3,388
linux-3.0	710K	13,856
ghostscript-9.00	1,363K	10,224

Challenge in Static Analysis

Precise, sound, scalable yet global static analyzers

Reality

Compromise either soundness or scalability

“bug-finders”	“verifiers”
scalable	sound
unsound	unscalable

Goal

Achieving scalable global static analyzers
without compromising precision and soundness

(http://www.spa-arrow.com)

Overall Approach

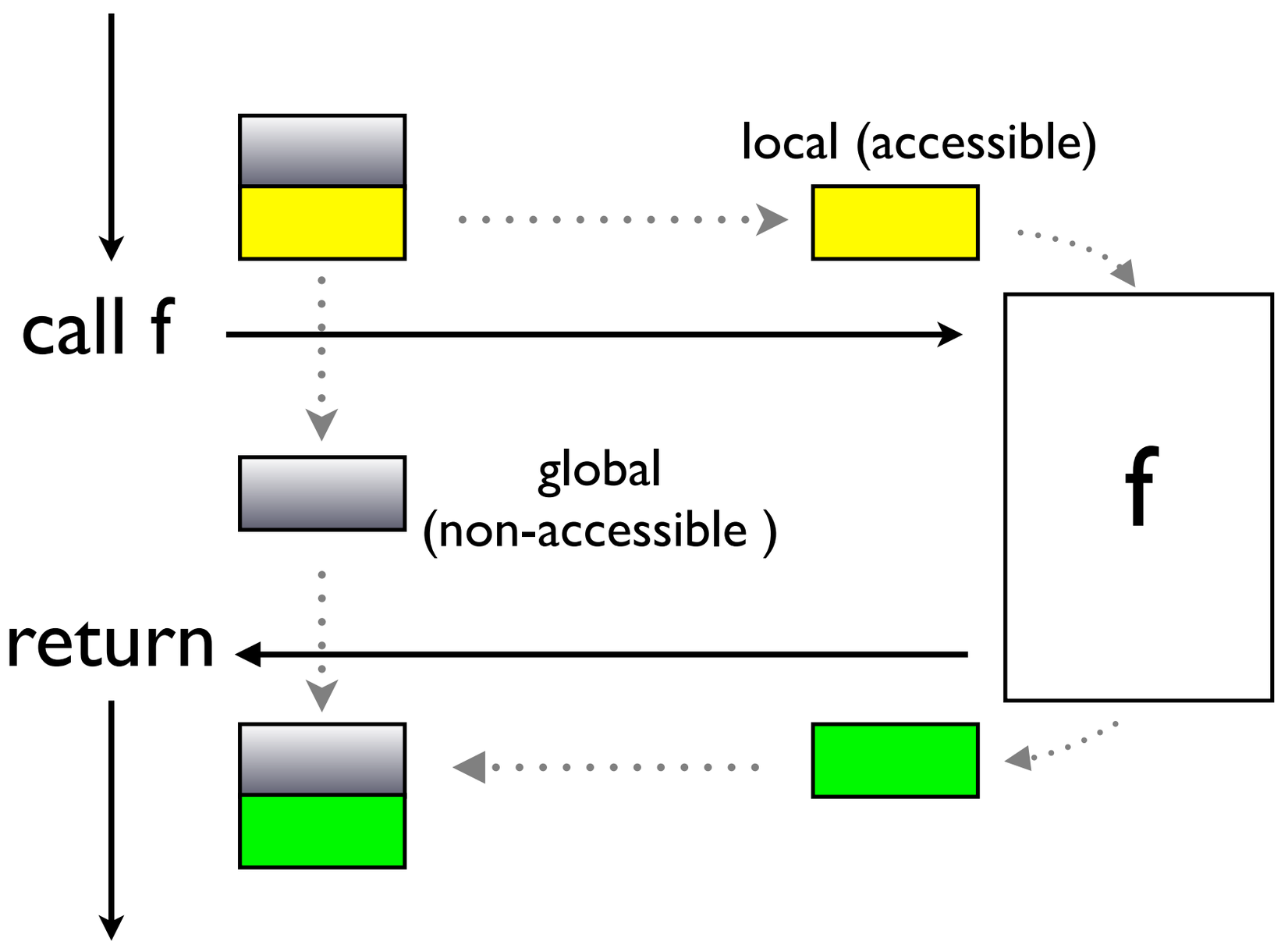
- Design by abstract interpretation
  - sound, precise, and global but unscalable
- Apply a set of cost-reduction techniques
  - scalable, preserving the precision and soundness

Localization

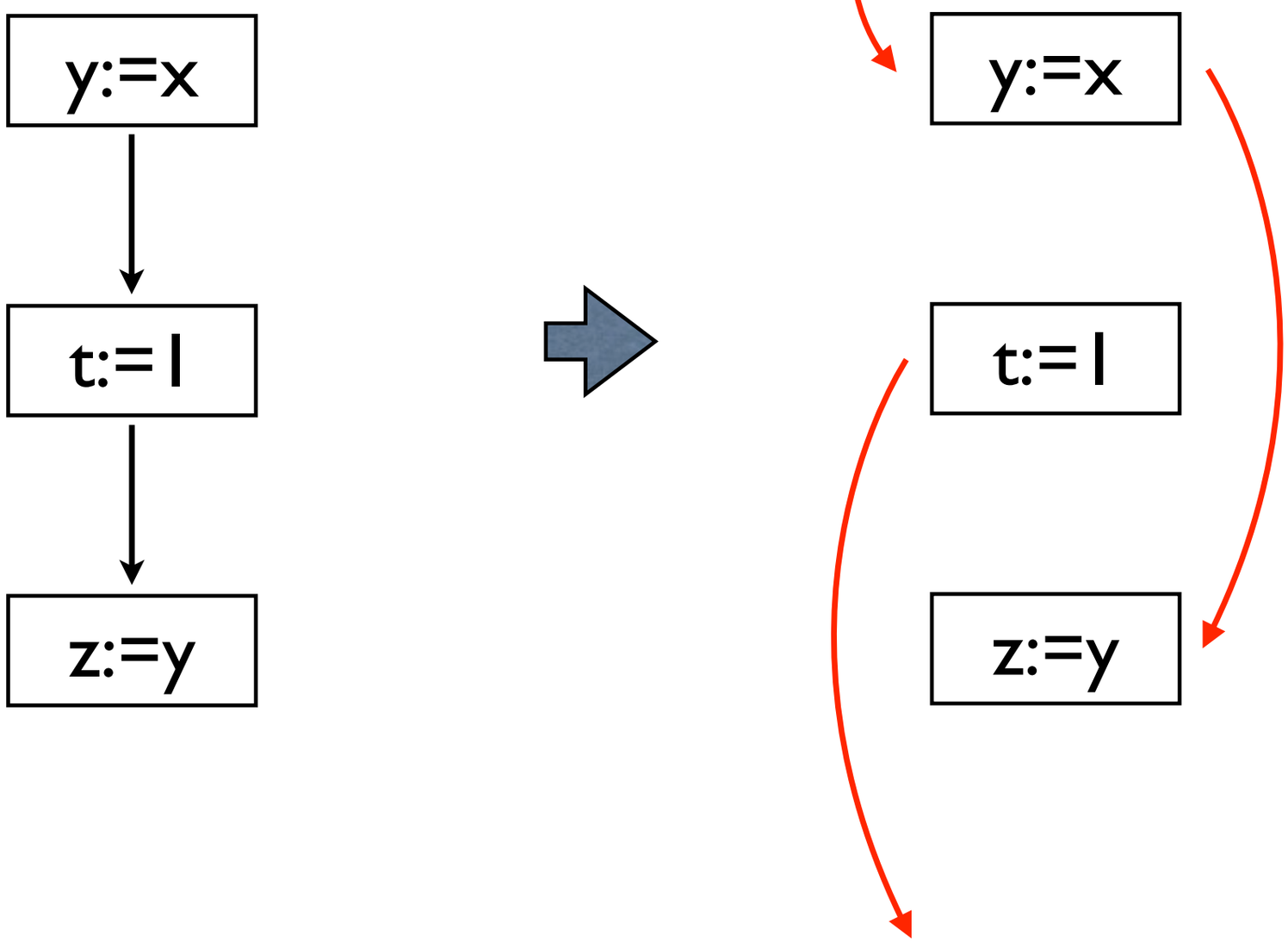
"local reasoning"
"framing" in separation logic

- **Spatial** localization
- **Temporal** localization

Spatial Localization



Temporal Localization



General Framework

"sparse analysis"

$$\hat{F} : \hat{D} \to \hat{D} \quad \Rightarrow \quad \hat{F}_s : \hat{D} \to \hat{D}$$

$$\vdots \qquad\qquad \vdots$$

$$\mathrm{lfp}\hat{F} \quad = \quad \mathrm{lfp}\hat{F}_s$$

Performance of Sparrow

The Early Bird

Program	LOC	Baseline		Localize		Spd↑	Mem↓
		Time	Mem	Time	Mem		
gzip-1.2.4a	7 K	772	240	3	63	257 x	74 %
bc-1.06	13 K	1,270	276	7	75	181 x	73 %
less-382	23 K	9,561	1,113	33	127	289 x	86 %
make-3.76.1	27 K	24,240	1,391	21	114	1,154 x	92 %
wget-1.9	35 K	44,092	2,546	11	85	4,008 x	97 %
a2ps-4.14	64 K	∞	N/A	40	353	N/A	N/A
sendmail-8.13.6	130 K	∞	N/A	744	678	N/A	N/A
nethack-3.3.0	211 K	∞	N/A	16,373	5,298	N/A	N/A
emacs-22.1	399 K	∞	N/A	37,830	7,795	N/A	N/A
python-2.5.1	435 K	∞	N/A	11,039	5,535	N/A	N/A
linux-3.0	710 K	∞	N/A	33,618	20,529	N/A	N/A
gimp-2.6	959 K	∞	N/A	3,874	3,602	N/A	N/A
ghostscript-9.00	1,363 K	∞	N/A	14,814	6,384	N/A	N/A

speed/memory much improved

scales to 1 million LOC

Summary

Scalable global static analysis is achievable

Sparse Analysis Framework

Thank you