THE ATTRIBUTOR: A VERSATILE INTER-PROCEDURAL FIXPOINT ITERATION FRAMEWORK

LLVM-Dev'19 — October 22, 2019 — San Jose, CA, USA

Johannes Doerfert*, Hideto Ueno, Stefan Stipanovic

*Leadership Computing Facility Argonne National Laboratory https://www.alcf.anl.gov/



Two of the authors were supported by Google Summer of Code (GSoC)!

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration) responsible for the planning and preparation of a capable exascale ecosystem, including software, applications, hardware, advanced system engineering, and early testbed platforms, in support of the nation's exascale computing imperative.

I. BACKGROUND

```
int * checkAndAdvance( int * __attribute__((aligned(16))) p ) {
    if (*p == 0)
        return checkAndAdvance(p + 4) ;
    return p ;
}
```



```
int * checkAndAdvance( int * __attribute__((aligned(16))) p ) {
    if (*p == 0)
        return checkAndAdvance(p + 4) ;
    return p;
}
```

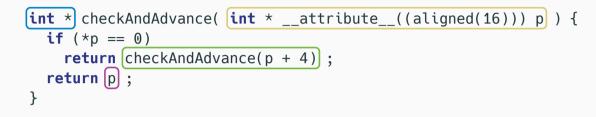
What is the alignment of:

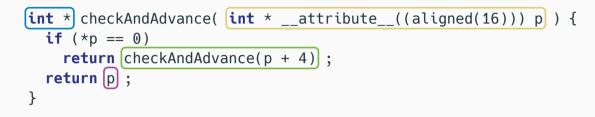
(1) the return type? (2) the returned value?

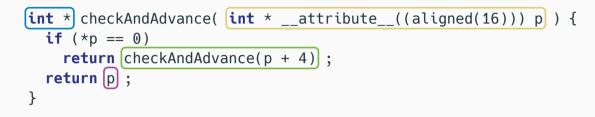
```
int * checkAndAdvance( int * __attribute__((aligned(16))) p ) {
    if (*p == 0)
        return checkAndAdvance(p + 4) ;
    return p;
}
```

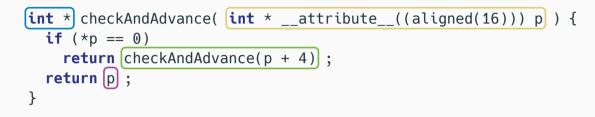
What is the alignment of:

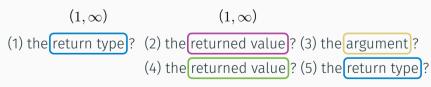
(1) the return type? (2) the returned value? (3) the argument?

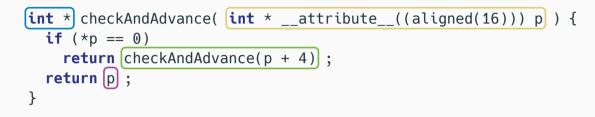




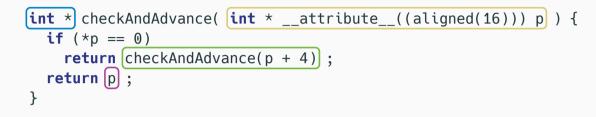


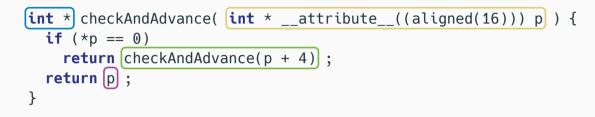




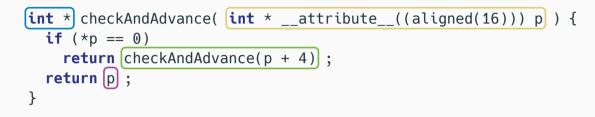


$$\begin{array}{ccc} (1,\infty) & (1,\infty) & (16,16) \\ (1) \mbox{ the return type} \mbox{?} & (2) \mbox{ the returned value} \mbox{?} & (3) \mbox{ the argument} \mbox{?} \\ & (4) \mbox{ the returned value} \mbox{?} & (5) \mbox{ the return type} \mbox{?} \end{array}$$



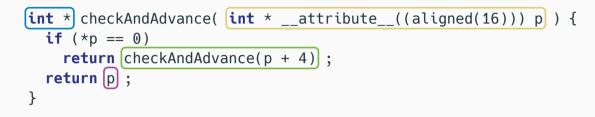


(1,
$$\infty$$
) (16, 16) \leftarrow (16, 16)
(1) the return type? (2) the returned value? (3) the argument?
(4) the returned value? (5) the return type?
(1, ∞)



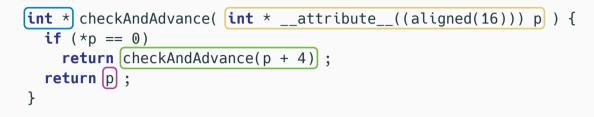
What is the alignment of:

(1, ∞) (16, 16) \leftarrow (16, 16) (1) the return type? (2) the returned value? (3) the argument? (4) the returned value? (5) the return type? (1, ∞) (1, ∞)



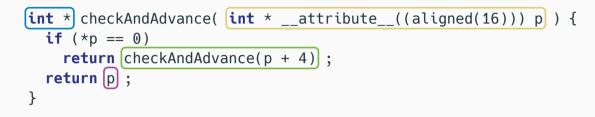
What is the alignment of:

(1, ∞) (16, 16) \leftarrow (16, 16) (1) the return type? (2) the returned value? (3) the argument? (4) the returned value? (5) the return type? (1, ∞) \leftarrow (1, ∞)



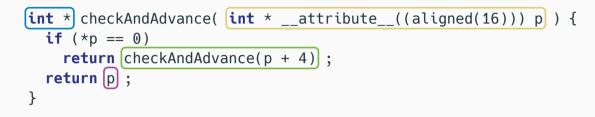
(1,16)
$$\leftarrow \cdots \odot \cdots \cdots (16,16) \leftarrow \cdots \cdots (16,16)$$

(1) the return type? (2) the returned value? (3) the argument?
(4) the returned value? (5) the return type?
(1, ∞) $\leftarrow \cdots \cdots (1, \infty)$



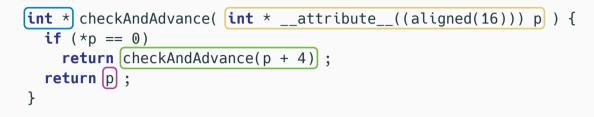
What is the alignment of:

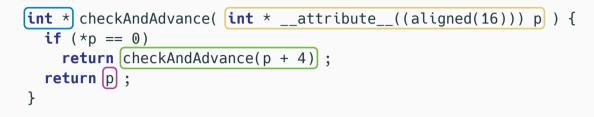
(1, 16) (16, 16) \leftarrow (16, 16) (1) the return type? (2) the returned value? (3) the argument? (4) the returned value? (5) the return type? (1, ∞) (1, 16)



What is the alignment of:

(1, 16) (16, 16) ← ----- (16, 16)
 (1) the return type? (2) the returned value? (3) the argument?
 (4) the returned value? (5) the return type?
 (1, 16) ← ----- (1, 16)

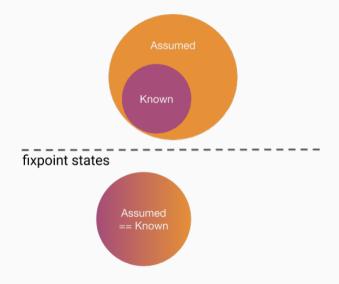




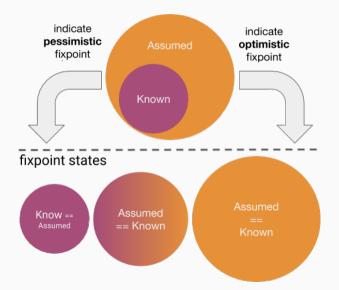
ABSTRACT STATES

ABSTRACT STATES





ABSTRACT STATES



```
int * checkAndAdvance( int * __attribute__((aligned(16))) p ) {
    if (*p == 0)
        return checkAndAdvance(p + 4) ;
    return p ;
}
```

Attributor A;

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn) ;
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

Attributor A;

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn);
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

Attributor A;

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn) ;
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

// Restrict deduction to specific abstract attributes.
auto Whitelist = {&AAAlign::ID};

Attributor A(Whitelist);

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn) ;
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

Attributor A(Whitelist);

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn) ;
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

AAAlign is unaware of AAIsDead and AAValueSimplify!

// Select what information is to be deduced.
IRPosition IRPRet = IRPosition::returned(Fn) ;
const auto &AA = A.getOrCreateAAFor< AAAlign >(IRPRet);

// Deduce information and manifest it in the IR.
auto Changed = A.run(*Fn->getParent());

At

THE ATTRIBUTOR — WHAT IT IS

- easy way to perform fixpoint analyses dependence tracking, work list algorithm, timeouts, ...

- easy way to perform fixpoint analyses dependence tracking, work list algorithm, timeouts, ...
- powerful way to perform fixpoint analyses utilize concurrently deduced information, e.g., liveness

- easy way to perform fixpoint analyses dependence tracking, work list algorithm, timeouts, ...
- powerful way to perform fixpoint analyses utilize concurrently deduced information, e.g., liveness
- alternative to inlining

IPO + internalization + function rewriting, e.g., argument promotion

- easy way to perform fixpoint analyses dependence tracking, work list algorithm, timeouts. ...



- alternative to inlining

IPO + internalization + function rewriting, e.g., argument promotion

II. MOTIVATION

THE ATTRIBUTOR — THE WHY IPO?

• recursion

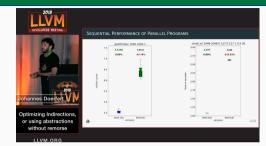
• recursion $\equiv loops$

- recursion \equiv *loops*
- code size

THE ATTRIBUTOR — THE WHY IPO?

inlining has limits:

- recursion $\equiv loops$
- code size

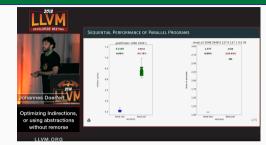


parallelism (think pthread_create) ↑

THE ATTRIBUTOR — THE WHY IPO?

inlining has limits:

- recursion \equiv *loops*
- code size



- parallelism (think pthread_create) ↑
- (declarations) \Rightarrow



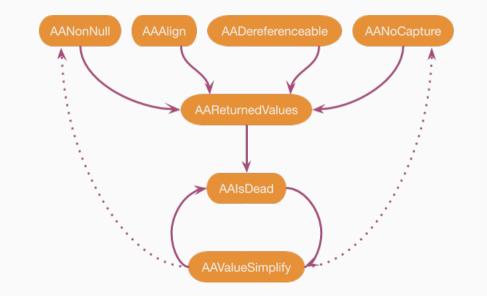
	MET CEAN, Argunna Hadismal Lak	
Writing Syllerizable Cede Is Hard	Introducing "Header Time Optimization"	Experiments
We does now the same the basic static for the point sector and the same fraction of the same static	A fit on of a simplifying proving, many sites with a distribution (SA with a d	Acceleration of a second secon
Radia amplification (a), (in a sp), and mental sequences are apprendent to the second sequences are apprendent apprendent to the second sequences and a second sequences are apprendent appre		$\begin{array}{c} \begin{array}{c} \\ \end{array} \\ \hline \end{array} \\ \hline \end{array} \\ \begin{array}{c} \\ \\ \\ \\ \end{array} \\ \end{array} \\ \begin{array}{c} \\ \end{array} \\ \begin{array}{c} \\ \\ \end{array} \\ \end{array} \\ \begin{array}{c} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \\ \end{array} \\ \end{array} \\ \end{array} \\ \end{array} \\ \begin{array}{c} \\ \end{array} \\ $
May no design an 1520 * Anning 20 June 30 405 JUJ a situation on mengin from, * E anyo on the autobiol Jupar Ankol, queening queen * E adapt separation on a 1500 apart on the proper to go uning an anneral May 35 Annotation and 1500 apart on the proper to go uning an anneral May 35 Annotation and 1600 apart of 450% queenings-ment how handle banklet May 36 Annotation and 1600 aparts and 1600 aparts and the site of the May 36 Annotation and 1600 aparts and 1600 aparts and the site of the May 36 Annotation and addition and the site of the site of the site of the site of the May 36 Annotation and addition and the site of the site		
Handler Flass Climate te service imposed the sequence of the comparison of the sequence of the	Appen Paper and the Kerry Ye and an even stated at the State State of the State Stat	Astropholyperaris & Feleranaes

Hashing plant development of initialized in any CACH program.
 Hashing a plant development of initialized in any CACH program.
 Hashing and plant development address in Cacity (Alama Manna).
 Hashing and the set of th

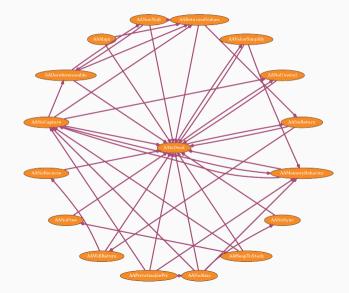
annine ((in, en) (minist)), argunnis, (ministration)) manna manifesta (in, ministrat) We is the event opportunity of the THE Apportunity for an even to be in the second of the HEER STATE o

Particular optical and an end of the second second

THE ATTRIBUTOR - WHY A FRAMEWORK?

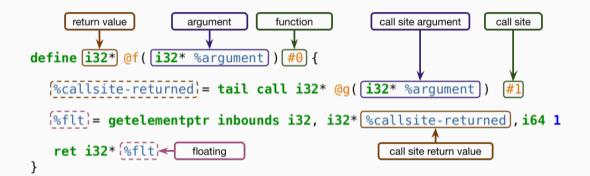


THE ATTRIBUTOR - WHY A FRAMEWORK?



III. DESIGN

LLVM-IR POSITIONS



AAVALUESIMPLIFYRETURNED::UPDATEIMPL(ATTRIBUTOR &A)

ChangeStatus updateImpl(Attributor &A) override {

```
Optional<Value *> After = getAssumedSimplifiedValue();
if (Before == After)
return ChangeStatus::UNCHANGED;
return ChangeStatus::CHANGED;
```

```
auto Pred = [&](Instruction &I) {
};
if (!A.checkForAllInstructions(Pred, this, {Instruction::Ret}))
return indicatePessimisticFixpoint();
```

```
Optional<Value *> After = getAssumedSimplifiedValue();
if (Before == After)
    return ChangeStatus::UNCHANGED;
return ChangeStatus::CHANGED;
```

```
Optional<Value *> After = getAssumedSimplifiedValue();
if (Before == After)
  return ChangeStatus::UNCHANGED;
return ChangeStatus::CHANGED;
```

```
auto Pred = [&](Instruction &I) {
    return combine(A.getAAFor<AAValueSimplify>(this, I.getOperand(0)));
};
if (!A.checkForAllInstructions(Pred, this, {Instruction::Ret}))
    return indicatePessimisticFixpoint();
```

```
Optional<Value *> After = getAssumedSimplifiedValue();
if (Before == After)
    return ChangeStatus::UNCHANGED;
return ChangeStatus::CHANGED;
```

NEW ATTRIBUTES

nofree

nosync

willreturn

dereferenceable_globally

NON-ATTRIBUTE DEDUCTIONS

liveness

returned values

value simplify

heap-2-stack

pointer privatization

THE ATTRIBUTOR — CHALLENGES

when to specialize for call sites $(\equiv$ "inlining + outlining")

how to seed abstract attributes (heuristics, pgo-based, ...)

reduce overheads

combine deduction schemes, e.g., context-based & def-use-based

THE ATTRIBUTOR — CHALLENGES

...

Evaluation — FunctionAttrs (late) vs. Attributor (early)

loc.	attribute	# w/o A.	# w/ A.	A. Δ	tot. w/o A.	tot. w/ A.
fn.	nosync	0	7612		0.0%	4.36%

EVALUATION — FUNCTIONATTRS (LATE) VS. ATTRIBUTOR (EARLY)

loc.	attribute	# w/o A.	# w/ A.	A. Δ	tot. w/o A.	tot. w/ A.
fn.	nosync	0	7612		0.0%	4.36%
arg. der	referenceable	61825	66317	+7.27%	35.4%	38.0%

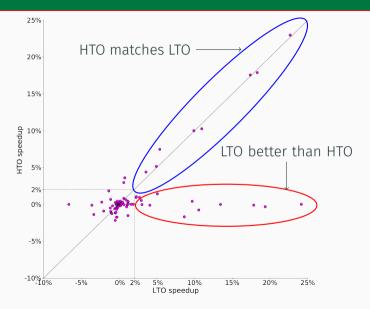
Evaluation — FunctionAttrs (late) vs. Attributor (early)

loc.	attribute	# w/o A.	# w/ A.	A. Δ	tot. w/o A.	tot. w/ A.
fn.	nosync	0	7612		0.0%	4.36%
arg.	dereferenceable	61825	66317	+7.27%	35.4%	38.0%
fn.	nofree	5762	10188	+76.81%	3.3%	5.83%
fn.	willreturn	0	4146		0.0%	2.37%
arg.	writeonly	0	3562		0.0%	2.04%
arg.	readnone	5377	6040	+12.33%	3.08%	3.46%
fn.	noreturn	965	1611	+66.94%	0.553%	0.923%
arg.	align	419	900	+114.80%	0.24%	0.515%
ret.	dereferenceable	19041	19479	+2.30%	11.2%	11.4%
arg.	nocapture	28991	29413	+1.46%	16.6%	16.8%
arg.	readonly	14946	15281	+2.24%	8.56%	8.75%
arg.	returned	512	599	+16.99%	0.293%	0.343%
arg.	noalias	4098	4158	+1.46%	2.35%	2.38%
ret.	noalias	1150	1194	+3.83%	0.676%	0.701%14

Evaluation — FunctionAttrs (late) vs. Attributor (early)

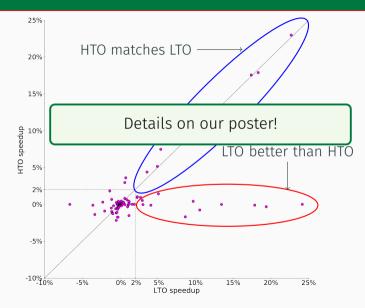
loc. attribute	# w/o A.	# w/ A.	A. Δ	tot. w/o A.	tot. w/ A.
fn. nosync	0	7612		0.0%	4.36%
arg. dereferenceable	61825	66317	+7.27%	35.4%	38.0%
fn. nofree	5762	10188	+76.81%	3.3%	5.83%
fn. willr gturn	0	1.11.6		0.0%	2.37%
arg. write	Details	Details on our poster!			2.04%
arg. read				3.08%	3.46%
fn. noreturn	965	1611	+66.94%	0.553%	0.923%
arg. align	419	900	+114.80%	0.24%	0.515%
ret. dereferenceable	19041	19479	+2.30%	11.2%	11.4%
arg. nocapture	28991	29413	+1.46%	16.6%	16.8%
arg. readonly	14946	15281	+2.24%	8.56%	8.75%
arg. returned	512	599	+16.99%	0.293%	0.343%
arg. noalias	4098	4158	+1.46%	2.35%	2.38%
ret. noalias	1150	1194	+3.83%	0.676%	0.701%14

Evaluation – (Attributor aided) "Header Time Optimiztion" (HTO)



15/16

Evaluation — (Attributor aided) "Header Time Optimiztion" (HTO)



1) introduce a **new llvm::Attribute**

introduce a new llvm::Attribute
 derive the new llvm::Attribute with the Attributor

1) introduce a **new llvm::Attribute**

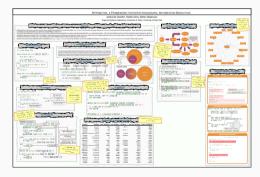
2) derive the new llvm::Attribute with the Attributor

3) use the **new llvm::Attribute** to improve alias analysis

Tutorial: tomorrow 1:45pm - 2:55pm **Posters**: tomorrow 4:00pm - 5:00pm

THE ATTRIBUTOR FRAMEWORK @ LLVM-Dev'19

Tutorial: tomorrow 1:45pm - 2:55pm Posters: tomorrow 4:00pm - 5:00pm



"Header Time Optimization": Cross-Translation Unit Optimization via Annotated Headers William 5. Moses farmoses limit edul. Johannes Doerfert lidoerfertiliant.cov

Writing Optimizable Code is Hav

The world by address remaining how most loss, pure division dyrages meterical engines, dyrages intering of models, ..., derivates (intering) ;

Bits & a problem increal programed in the ECO EXMercit involvement [2] milling reset? In fact, page gives a PE improvement in the restre program both another

Header Films

Introducing 'Header Time Optimizatio

4 stars themptotics





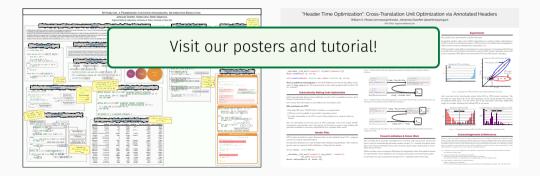
Present Limitations & Future Work





THE ATTRIBUTOR FRAMEWORK @ LLVM-Dev'19

Tutorial: tomorrow 1:45pm - 2:55pm **Posters**: tomorrow 4:00pm - 5:00pm



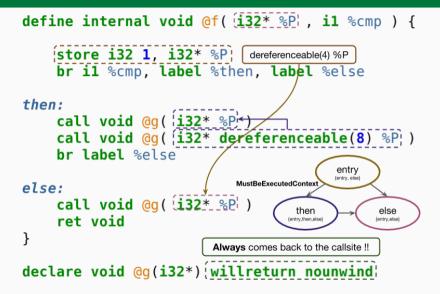


THE ATTRIBUTOR — EVALUATION — ASSUMING EXACT DEFINITIONS

attribute	# w/o A.	# w/ A.	A. Δ	tot. w/o A.	tot. w/ A.
nosync	0	78491		0.0%	45.90%
dereferenceable	59578	64214	+7.78%	34.8%	37.50%
nofree	25649	76719	+199.11%	15.0%	44.90%
willreturn	0	64748		0.0%	37.90%
writeonly	0	4229		0.0%	2.47%
readnone	40505	38414	-5.16%	23.7%	22.50%
noreturn	879	2394	+172.36%	0.514%	1.40%
align	449	1028	+128.95%	0.263%	0.60%
dereferenceable	18064	19419	+7.50%	10.8%	11.60%
nocapture	153523	155294	+1.15%	89.8%	90.80%
returned	9418	13937	+47.98%	5.51%	8.15%
noalias	4113	4189	+1.85%	2.41%	2.45%
noalias	3015	3310	+9.78%	1.81%	1.98%
writeonly	8089	9877	+22.10%	4.73%	5.78%
nounwind	123516	125480	+1.59%	72.2%	73.40%
	nosync dereferenceable nofree willreturn writeonly readnone noreturn align dereferenceable nocapture returned noalias noalias writeonly	nosync0dereferenceable59578nofree25649willreturn0writeonly0readnone40505noreturn879align449dereferenceable18064nocapture153523returned9418noalias4113noalias3015writeonly8089	nosync078491dereferenceable5957864214nofree2564976719willreturn064748writeonly04229readnone4050538414noreturn8792394align4491028dereferenceable1806419419nocapture153523155294returned941813937noalias41134189noalias30153310writeonly80899877	nosync078491dereferenceable5957864214+7.78%nofree2564976719+199.11%willreturn064748writeonly04229readnone4050538414-5.16%noreturn8792394+172.36%align4491028+128.95%dereferenceable1806419419+7.50%nocapture153523155294+1.15%returned941813937+47.98%noalias41134189+1.85%noalias30153310+9.78%writeonly80899877+22.10%	nosync0784910.0%dereferenceable5957864214+7.78%34.8%nofree2564976719+199.11%15.0%willreturn0647480.0%writeonly042290.0%readnone4050538414-5.16%23.7%noreturn8792394+172.36%0.514%align4491028+128.95%0.263%dereferenceable1806419419+7.50%10.8%nocapture153523155294+1.15%89.8%returned941813937+47.98%5.51%noalias41134189+1.85%2.41%noalias30153310+9.78%1.81%writeonly80899877+22.10%4.73%

MUST-BE-EXECUTED-CONTEXT

MUST-BE-EXECUTED-CONTEXT



INLINING VS. IPO

The "inline-first" approach:

- I: aggressive inlining, e.g., all N call sites
- **II:** perform intra-procedural analyses + transformations (*N* times)
- III: derive information + transformation opportunities inter-procedurally

The "inline-first" approach:

- I: aggressive inlining, e.g., all N call sites
- **II:** perform intra-procedural analyses + transformations (*N* times)
- III: derive information + transformation opportunities inter-procedurally

The *"IPO-first"* approach:

I: derive information + transformation opportunities inter-procedurally
II: internalize & specialize functions if necessary & beneficial
III: inline where benefit can be expected