

A Theory of Arrays with set and copy Operations

Stephan Falke, Carsten Sinz, and Florian Merz

INSTITUTE FOR THEORETICAL COMPUTER SCIENCE (ITI)

```
struct list_node {  
    int data;  
    struct list_node *tail;  
};  
typedef struct list_node list;  
  
list *reverse(list *l) {  
    list *r = l, *p = NULL;  
    while (r != NULL) {  
        list *q = r;  
        r = r->tail;  
        q->tail = p;  
        p = q;  
    }  
    return p;  
}
```



- SMT-solvers are routinely used in program analysis:
 - Deductive program verification
 - Symbolic execution
 - Bounded model checking
 - ...

- SMT-solvers are routinely used in program analysis:
 - Deductive program verification
 - Symbolic execution
 - Bounded model checking
 - ...
- Prominent theory: \mathcal{T}_A (theory of arrays)
 - Model arrays/structures/objects in the program
 - Model main memory

What to do with standard library functions such as `memset` and `memcpy/memmove`?

```
void *memset(void *s, int c, size_t n);
```

```
void *memcpy(void *dest, const void *src, size_t n);
```

```
void *memmove(void *dest, const void *src, size_t n);
```

What to do with standard library functions such as `memset` and `memcpy/memmove`?

might not be constant-sized!

```
void *memset(void *s, int c, size_t n);
```

might not be constant-sized!

```
void *memcpy(void *dest, const void *src, size_t n);
```

might not be constant-sized!

```
void *memmove(void *dest, const void *src, size_t n);
```

What to do with standard library functions such as `memset` and `memcpy/memmove`?

might not be constant-sized!

```
void *memset(void *s, int c, size_t n);
```

might not be constant-sized!

```
void *memcpy(void *dest, const void *src, size_t n);
```

might not be constant-sized!

```
void *memmove(void *dest, const void *src, size_t n);
```

⇒ Extend \mathcal{T}_A to allow for **dynamically-sized read/writes!**

- 1 \mathcal{T}_{ASC} : an extension of \mathcal{T}_A with set and copy operations
 - set: sets ranges in an array at once
 - copy: copies ranges between arrays

- 1 \mathcal{T}_{ASC} : an extension of \mathcal{T}_A with set and copy operations
 - set: sets ranges in an array at once
 - copy: copies ranges between arrays
- 2 Satisfiability checking for \mathcal{T}_{ASC} by reducing to \mathcal{T}_A + index arithmetic

- Precisely model `memset` and `memcpy/memmove`

Applications of \mathcal{T}_{ASC}

- Precisely model `memset` and `memcpy/memmove`
- Zero initialization of global variables

Applications of \mathcal{T}_{ASC}

- Precisely model `memset` and `memcpy/memmove`
- Zero initialization of global variables
- Zero initialization of fresh memory pages

- Precisely model `memset` and `memcpy/memmove`
- Zero initialization of global variables
- Zero initialization of fresh memory pages
- “Havoc” memory regions (volatile variables)

Applications of \mathcal{T}_{ASC}

- Precisely model `memset` and `memcpy/memmove`
- Zero initialization of global variables
- Zero initialization of fresh memory pages
- “Havoc” memory regions (volatile variables)
- Model memory mapped I/O

- Precisely model `memset` and `memcpy/memmove`
- Zero initialization of global variables
- Zero initialization of fresh memory pages
- “Havoc” memory regions (volatile variables)
- Model memory mapped I/O
- Attaching metadata to memory regions (allocated, de-allocated, ...)

Sorts	E : elements I : indices A : arrays
Functions	read : $A \times I \rightarrow E$ write : $A \times I \times E \rightarrow A$


$$p = r \quad \Longrightarrow \quad \text{read}(\text{write}(a, p, v), r) = v$$

$$\neg(p = r) \quad \Longrightarrow \quad \text{read}(\text{write}(a, p, v), r) = \text{read}(a, r)$$

\mathcal{T}_A : The Theory of Arrays

Sorts	E : elements I : indices A : arrays
Functions	read : $A \times I \rightarrow E$ write : $A \times I \times E \rightarrow A$

a write modifies the position written to ...


 $p = r \implies \text{read}(\text{write}(a, p, v), r) = v$


$$\neg(p = r) \implies \text{read}(\text{write}(a, p, v), r) = \text{read}(a, r)$$

\mathcal{T}_A : The Theory of Arrays

Sorts	E : elements I : indices A : arrays
Functions	read : $A \times I \rightarrow E$ write : $A \times I \times E \rightarrow A$

a write modifies the position written to ...

 $p = r \implies \text{read}(\text{write}(a, p, v), r) = v$

 $\neg(p = r) \implies \text{read}(\text{write}(a, p, v), r) = \text{read}(a, r)$

...and nothing else.

\mathcal{T}_A : The Theory of Arrays

Sorts	E : elements I : indices A : arrays
Functions	read : $A \times I \rightarrow E$ write : $A \times I \times E \rightarrow A$

a write modifies the position written to ...

\curvearrowright $p = r \implies \text{read}(\text{write}(a, p, v), r) = v$

$\curvearrowright \neg(p = r) \implies \text{read}(\text{write}(a, p, v), r) = \text{read}(a, r)$

...and nothing else.

replace $\text{read}(\text{write}(a, p, v), r)$ by $\text{ITE}(p = r, v, \text{read}(a, r))$
treat remaining arrays as uninterpreted functions

The Theory of Arrays with `set` and `copy`

Sorts	$S = I$
Functions	$\text{set} : A \times I \times E \times S \rightarrow A$ $\text{copy} : A \times I \times A \times I \times S \rightarrow A$

$$p \leq r < p + s \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\neg(p \leq r < p + s) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

The Theory of Arrays with `set` and `copy`

Sorts	$S = I$
Functions	$\text{set} : A \times I \times E \times S \rightarrow A$ $\text{copy} : A \times I \times A \times I \times S \rightarrow A$


$$S = I = \mathbb{N}$$

index type needs arithmetic

$$\underbrace{p \leq r < p + s} \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\neg(p \leq r < p + s) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

The Theory of Arrays with `set` and `copy`

Sorts	$S = I$
Functions	$\text{set} : A \times I \times E \times S \rightarrow A$ $\text{copy} : A \times I \times A \times I \times S \rightarrow A$

$S = I = \mathbb{N}$

index type needs arithmetic

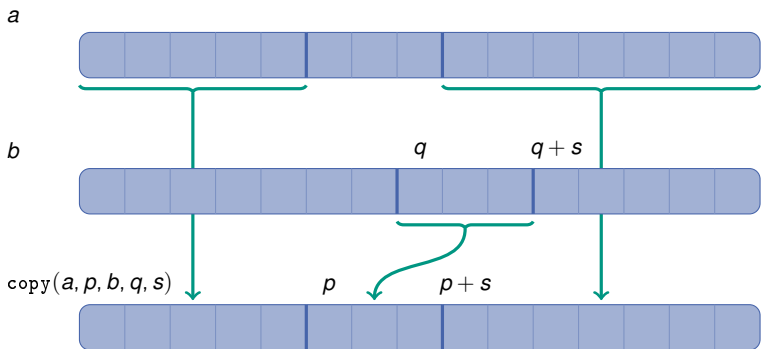
$$\underbrace{p \leq r < p + s} \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\neg(p \leq r < p + s) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

$$p \leq r < p + s \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(b, q + r - p)$$

$$\neg(p \leq r < p + s) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(a, r)$$

The copy Operation: Illustration



Sorts	$S = I$
Functions	$\text{set}_\infty : A \times I \times E \rightarrow A$ $\text{copy}_\infty : A \times I \times A \times I \rightarrow A$ $\text{set} : A \times I \times E \times S \rightarrow A$ $\text{copy} : A \times I \times A \times I \times S \rightarrow A$

$$p \leq r \implies \text{read}(\text{set}_\infty(a, p, v), r) = v$$

$$\neg(p \leq r) \implies \text{read}(\text{set}_\infty(a, p, v), r) = \text{read}(a, r)$$

$$p \leq r \implies \text{read}(\text{copy}_\infty(a, p, b, q), r) = \text{read}(b, q + r - p)$$

$$\neg(p \leq r) \implies \text{read}(\text{copy}_\infty(a, p, b, q), r) = \text{read}(a, r)$$

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)) \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \underbrace{\text{bvadd}(p, s)}_{\text{might overflow}}) \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \underbrace{\text{bvadd}(p, s)}_{\text{might overflow}})) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

$$\underbrace{\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))}_{\text{always false on overflow}} \implies \text{read}(\text{set}(a, p, v, s), r) = v$$

might overflow

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

might overflow

$$\underbrace{\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))}_{\text{always false on overflow}} \stackrel{\text{might overflow}}{\implies} \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\underbrace{\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)))}_{\text{might overflow}} \stackrel{\text{always true on overflow}}{\implies} \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

$$\underbrace{\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))}_{\text{always false on overflow}} \overset{\text{might overflow}}{\implies} \text{read}(\text{set}(a, p, v, s), r) = v$$

$$\underbrace{\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)))}_{\text{might overflow}} \overset{\text{always true on overflow}}{\implies} \text{read}(\text{set}(a, p, v, s), r) = \text{read}(a, r)$$

\implies If $\text{bvadd}(p, s)$ overflows, then $\text{set}(a, p, v, s)$ is a **no-op**

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(b, \text{bvadd}(q, \text{bvsub}(r, p)))$$

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(a, r)$$

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(b, \text{bvadd}(q, \text{bvsb}(r, p)))$$

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(a, r)$$

\implies If $\text{bvadd}(p, s)$ overflows, then $\text{copy}(a, p, b, q, s)$ is a **no-op**

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(b, \text{bvadd}(q, \underbrace{\text{bvsub}(r, p)}_{\text{underflow guarded}}))$$

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(a, r)$$

\implies If $\text{bvadd}(p, s)$ overflows, then $\text{copy}(a, p, b, q, s)$ is a **no-op**

$$\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s)) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(b, \text{bvadd}(q, \text{bvsub}(r, p)))$$

underflow guarded
range wraps around on overflow

$$\neg(\text{bvule}(p, r) \wedge \text{bvult}(r, \text{bvadd}(p, s))) \implies \text{read}(\text{copy}(a, p, b, q, s), r) = \text{read}(a, r)$$

\implies If $\text{bvadd}(p, s)$ overflows, then $\text{copy}(a, p, b, q, s)$ is a **no-op**

- Based on **reductions** to \mathcal{T}_A + index arithmetic

- Based on **reductions** to \mathcal{T}_A + index arithmetic
- 3 **quantifier-based** approaches

- Based on **reductions** to \mathcal{T}_A + index arithmetic
- 3 **quantifier-based** approaches
- 2 **quantifier-free** approaches
 - Eager encoding
 - Instantiation-based approach

■ Fully Quantified Axioms

$$\forall a, p, v, s, r. \text{read}(\text{set}(a, p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Fully Quantified Axioms

$$\forall a, p, v, s, r. \text{read}(\text{set}(a, p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Quantified Axioms Instantiated for Arrays

$$\forall p, v, s, r. \text{read}(\text{set}_a(p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Fully Quantified Axioms

$$\forall a, p, v, s, r. \text{read}(\text{set}(a, p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Quantified Axioms Instantiated for Arrays

$$\forall p, v, s, r. \text{read}(\text{set}_a(p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Quantified Axioms Instantiated for all Arguments

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \\ \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

■ Fully Quantified Axioms

$$\forall a, p, v, s, r. \text{read}(\text{set}(a, p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Quantified Axioms Instantiated for Arrays

$$\forall p, v, s, r. \text{read}(\text{set}_a(p, v, s), r) = \\ \text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$$

■ Quantified Axioms Instantiated for all Arguments

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \\ \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

■ Require a solver that supports **quantifiers**

- Replace $\text{read}(\text{set}(a, p, v, s), r)$ by $\text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$

- Replace $\text{read}(\text{set}(a, p, v, s), r)$ by $\text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$
- \mathcal{T}_{ASC} axioms are **applied** eagerly

- Replace $\text{read}(\text{set}(a, p, v, s), r)$ by $\text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$
- \mathcal{T}_{ASC} axioms are **applied** eagerly
- Requires eager application of \mathcal{T}_A 's read-over-write axioms

- Replace $\text{read}(\text{set}(a, p, v, s), r)$ by $\text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$
- \mathcal{T}_{ASC} axioms are **applied** eagerly
- Requires eager application of \mathcal{T}_A 's read-over-write axioms
- Can be implemented as a **pre-processing step**

- Replace $\text{read}(\text{set}(a, p, v, s), r)$ by $\text{ITE}(p \leq r < p + s, v, \text{read}(a, r))$
- \mathcal{T}_{ASC} axioms are **applied** eagerly
- Requires eager application of \mathcal{T}_A 's read-over-write axioms
- Can be implemented as a **pre-processing step**
- Can be used in combination with **any** solver for \mathcal{T}_A + index arithmetic

Instantiation-Based Approach

- Replace $\text{set}(a, p, v, s)$ by a fresh constant symbol

- Replace $\text{set}(a, p, v, s)$ by a fresh constant symbol
- Needed **instantiations** of

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

are **added** eagerly

- Replace $\text{set}(a, p, v, s)$ by a fresh constant symbol
- Needed **instantiations** of

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

are **added** eagerly

- **One step beyond** Quantified Axioms Instantiated for all Arguments

- Replace $\text{set}(a, p, v, s)$ by a fresh constant symbol
- Needed **instantiations** of

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

are **added** eagerly

- **One step beyond** Quantified Axioms Instantiated for all Arguments
- Can be implemented as a **pre-processing step**

- Replace $\text{set}(a, p, v, s)$ by a fresh constant symbol
- Needed **instantiations** of

$$\forall r. \text{read}(\text{set}_{a,p,v,s}, r) = \text{ITE}(p' \leq r < p + s, v, \text{read}(a, r))$$

are **added** eagerly

- **One step beyond** Quantified Axioms Instantiated for all Arguments
- Can be implemented as a **pre-processing step**
- Can be used in combination with **any** solver for \mathcal{T}_A + index arithmetic

- Done in the software bounded model checker [LLBMC](#)

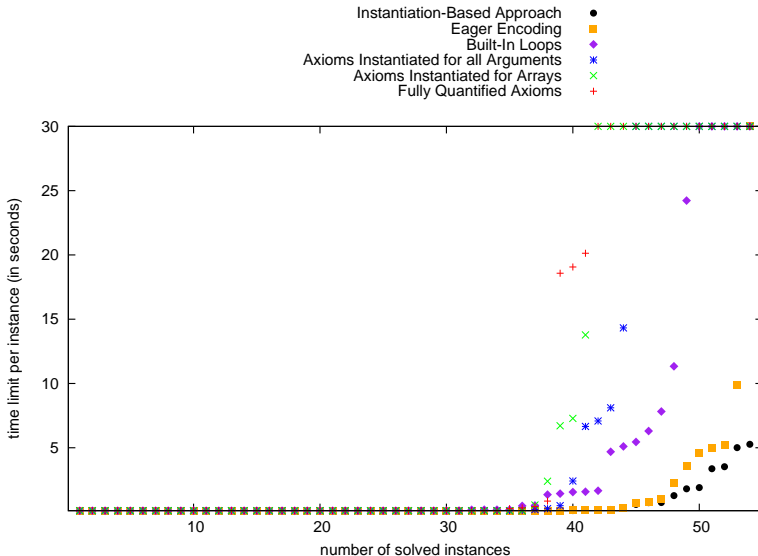
Evaluation

- Done in the software bounded model checker [LLBMC](#)
- Uses [bitvectors](#) as index type

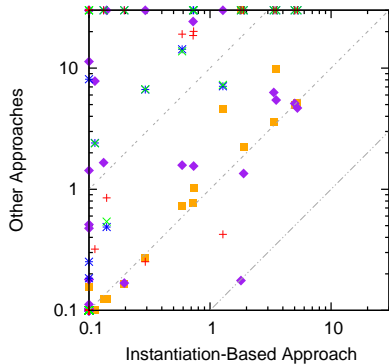
- Done in the software bounded model checker **LLBMC**
- Uses **bitvectors** as index type
- Applied on 55 examples
 - 19 programs from LLBMC's own \mathcal{T}_{ASC} test suite
 - 19 programs from LLBMC's own pre- \mathcal{T}_{ASC} test suite
 - 16 programs from the SMT-CBMC software bounded model checker
 - 1 program from the KLEE symbolic execution tool






- Done in the software bounded model checker **LLBMC**
- Uses **bitvectors** as index type
- Applied on 55 examples
 - 19 programs from LLBMC's own \mathcal{T}_{ASC} test suite
 - 19 programs from LLBMC's own pre- \mathcal{T}_{ASC} test suite
 - 16 programs from the SMT-CBMC software bounded model checker
 - 1 program from the KLEE symbolic execution tool
- Evaluated all 5 encodings + built-in loops
 - Quantifier-based approaches used Z3
 - Eager- and instantiation-based approaches used STP
 - Built-in loops approach used STP

Results



Results



Eager Encoding 
Built-In Loops 
Axioms Instantiated for all Arguments 
Axioms Instantiated for Arrays 
Fully Quantified Axioms 

Conclusion and Future Work

- \mathcal{T}_{ASC} is an **useful extension** of \mathcal{T}_A

Conclusion and Future Work

- \mathcal{T}_{ASC} is an **useful extension** of \mathcal{T}_A
- Performs **better than unrolling** loops for `memset` and `memcpy`

- \mathcal{T}_{ASC} is an **useful extension** of \mathcal{T}_A
- Performs **better than unrolling** loops for `memset` and `memcpy`
- We can do **better than z3's reasoning** involving **quantifiers**

- \mathcal{T}_{ASC} is an **useful extension** of \mathcal{T}_A
- Performs **better than unrolling** loops for `memset` and `memcpy`
- We can do **better than Z3's** reasoning involving **quantifiers**
- “**Lemmas-on-demand**”/“**lazy instantiation**” is the next step

`http://llbmc.org`