



Constraint Solving in DSE:

THE GOOD THE BAD AND THE UGLY

Andrea Mattavelli

Department of Computing
Imperial College London

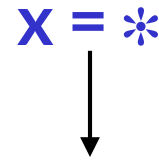
Dynamic Symbolic Execution

```
int bad_abs(int x) {  
    if (x < 0)  
        return -x;  
    if (x == 1234)  
        return -x;  
    return x;  
}
```

Dynamic Symbolic Execution

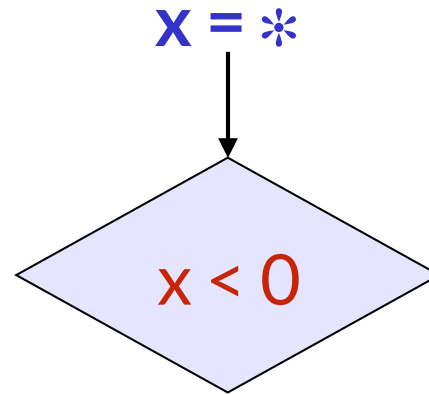
```
int bad_abs(int x) {  
    if (x < 0)  
        return -x;  
    if (x == 1234)  
        return -x;  
    return x;  
}
```

x = *



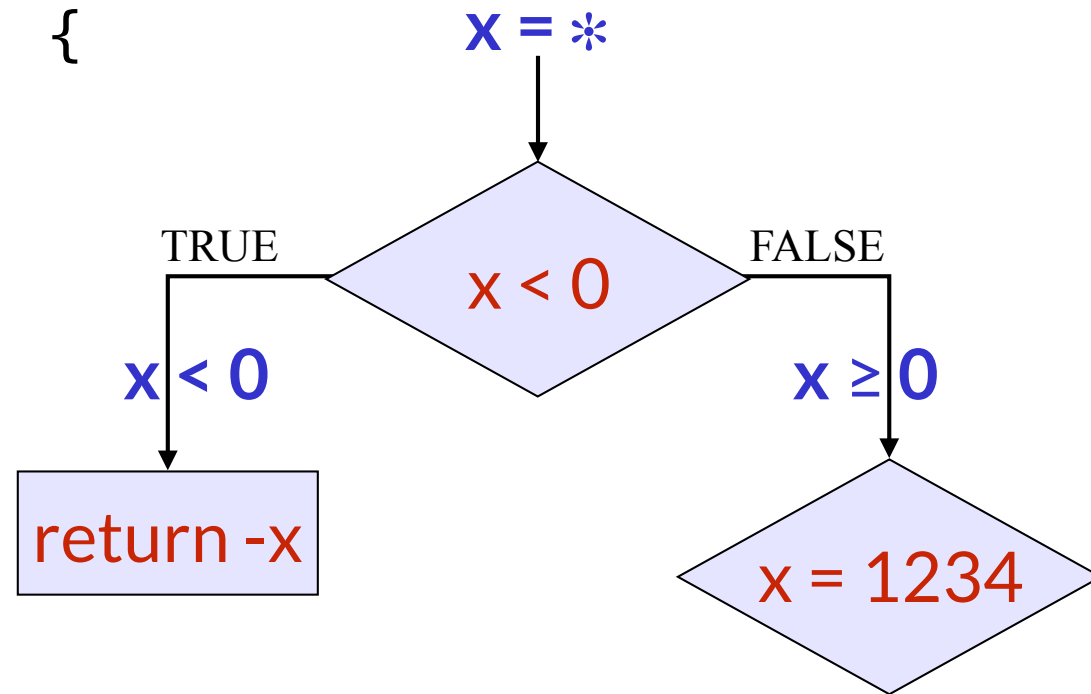
Dynamic Symbolic Execution

```
int bad_abs(int x) {  
  if (x < 0)  
    return -x;  
  if (x == 1234)  
    return -x;  
  return x;  
}
```



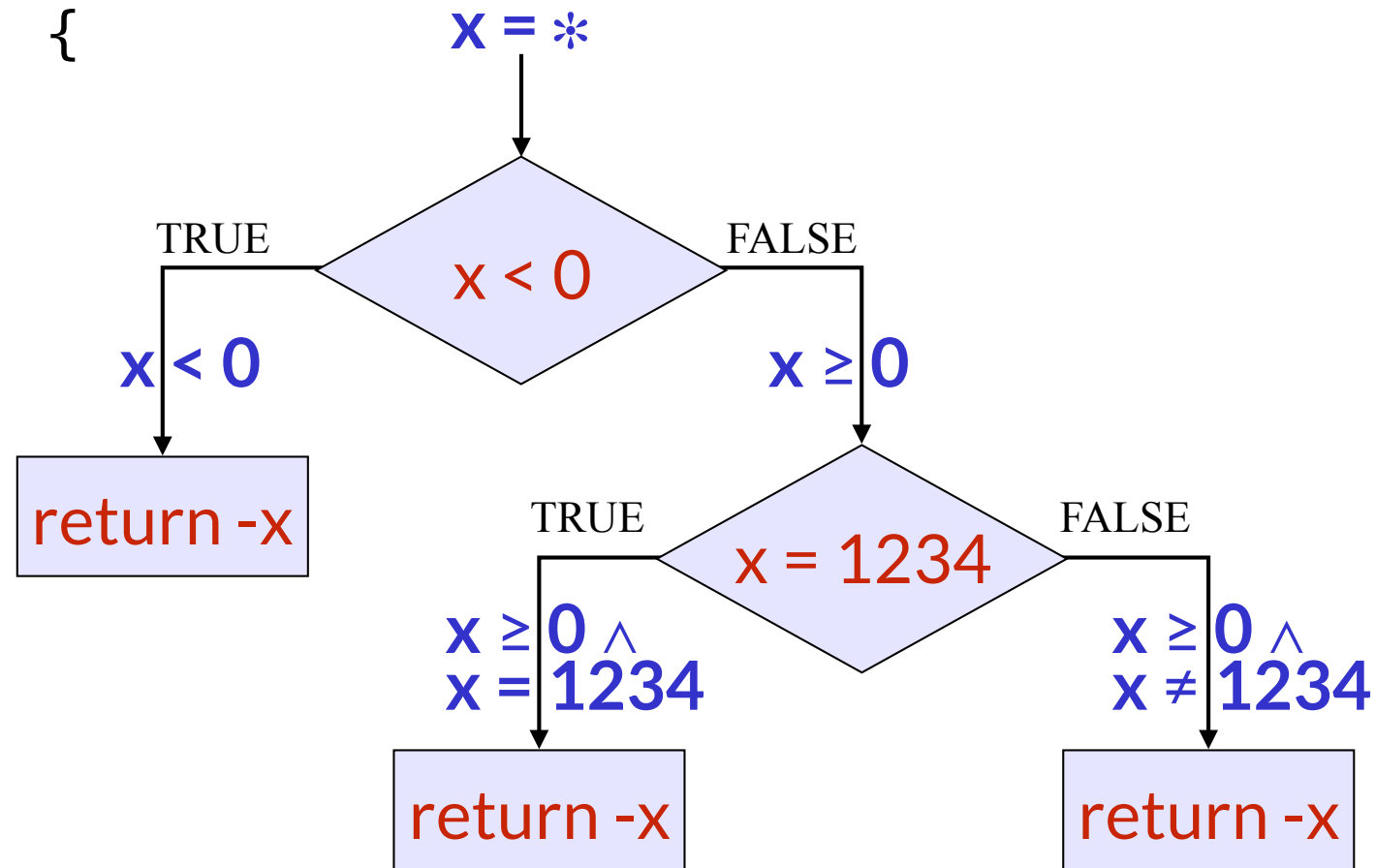
Dynamic Symbolic Execution

```
int bad_abs(int x) {  
    if (x < 0)  
        return -x;  
    if (x == 1234)  
        return -x;  
    return x;  
}
```



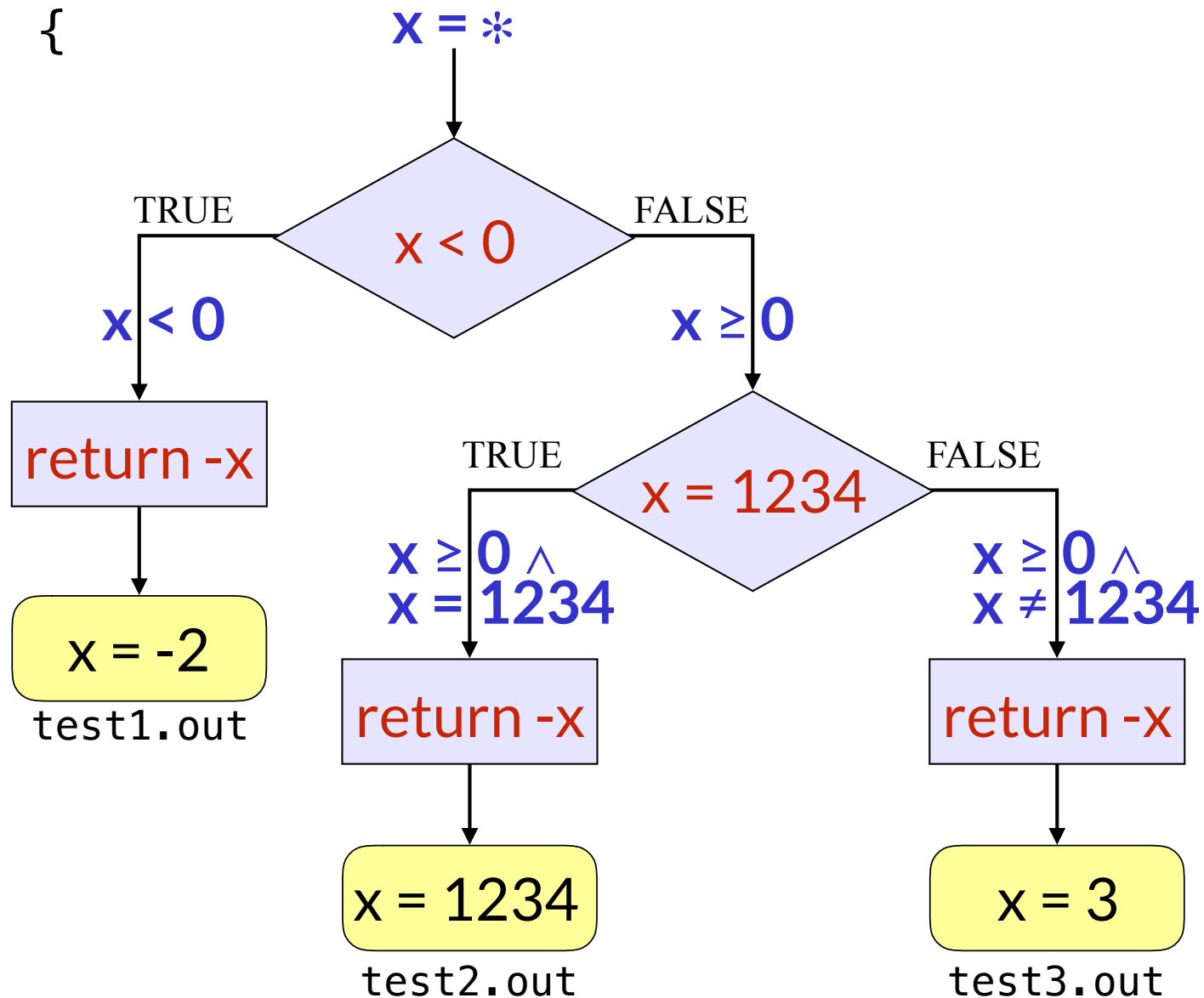
Dynamic Symbolic Execution

```
int bad_abs(int x) {  
  if (x < 0)  
    return -x;  
  if (x == 1234)  
    return -x;  
  return x;  
}
```



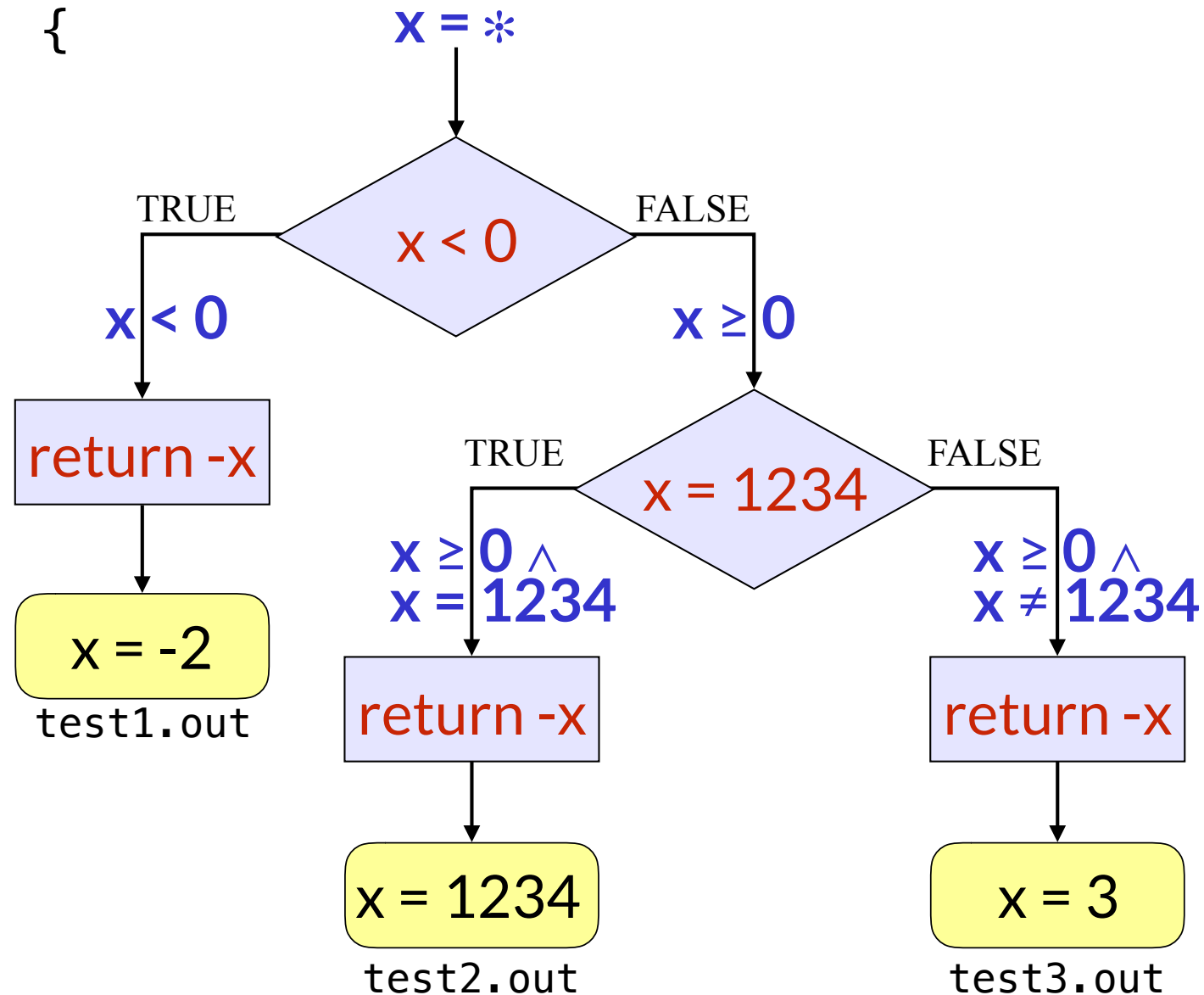
Dynamic Symbolic Execution

```
int bad_abs(int x) {  
  if (x < 0)  
    return -x;  
  if (x == 1234)  
    return -x;  
  return x;  
}
```



Constraint Solving & DSE

```
int bad_abs(int x) {  
  if (x < 0)  
    return -x;  
  if (x == 1234)  
    return -x;  
  return x;  
}
```



The Good: Checks and Accuracy



The Good: All-Value Checks

Implicit checks before each dangerous operation

- Null-pointer dereferences
- Buffer overflows
- Division/modulo by zero
- Asserts violations

All-value checks!

Errors are found if **any** buggy value exists on that path!

The Good: All-Value Checks

Implicit checks before each dangerous operation

- Null-pointer dereferences
- Buffer overflows
- Division/modulo by zero
- Asserts violations

All-value checks!

Errors are found if **any** buggy value exists on that path!

```
int foo(unsigned k) {
    int a[4] = {3, 1, 0, 4};
    k = k % 4;
    return a[a[k]];
}
```

The Good: All-Value Checks

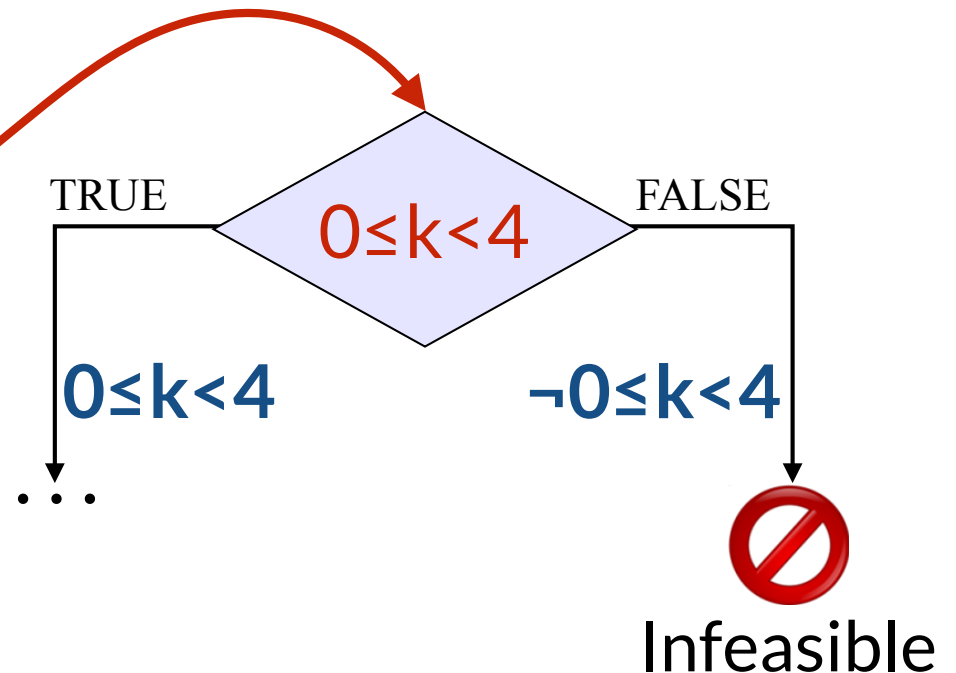
Implicit checks before each dangerous operation

- Null-pointer dereferences
- Buffer overflows
- Division/modulo by zero
- Asserts violations

All-value checks!

Errors are found if **any** buggy value exists on that path!

```
int foo(unsigned k) {  
    int a[4] = {3, 1, 0, 4};  
    k = k % 4;  
    return a[a[k]];  
}
```



The Good: All-Value Checks

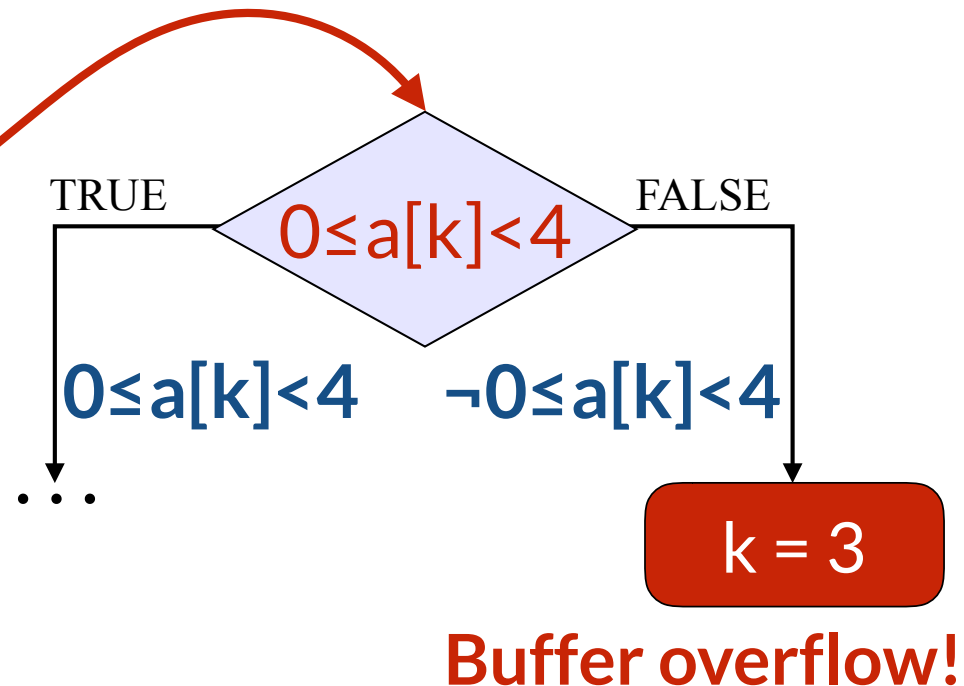
Implicit checks before each dangerous operation

- Null-pointer dereferences
- Buffer overflows
- Division/modulo by zero
- Asserts violations

All-value checks!

Errors are found if **any** buggy value exists on that path!

```
int foo(unsigned k) {  
    int a[4] = {3, 1, 0, 4};  
    k = k % 4;  
    return a[a[k]];  
}
```



The Good: Accuracy

Need constraint solver that allows bit-level modeling of memory:

- Systems code often observes the **same bytes** in **different ways**
- Bugs in systems code are often triggered by **corner cases**

The Good: Accuracy

Need constraint solver that allows bit-level modeling of memory:

- Systems code often observes the **same bytes** in **different ways**
- Bugs in systems code are often triggered by **corner cases**

```
char buf[N]; // symbolic
struct pkt1 { char x,y,v,w; int z; } *pa = (struct pkt1*) buf;
struct pkt2 { unsigned i, j; } *pb = (struct pkt2*) buf;
if (pa[2].v < 0) { assert(pb[2].i >= 1<<23); }
```

The Good: Accuracy

Need constraint solver that allows bit-level modeling of memory:

- Systems code often observes the **same bytes** in **different ways**
- Bugs in systems code are often triggered by **corner cases**

```
char buf[N]; // symbolic
struct pkt1 { char x,y,v,w; int z; } *pa = (struct pkt1*) buf;
struct pkt2 { unsigned i, j; } *pb = (struct pkt2*) buf;
if (pa[2].v < 0) { assert(pb[2].i >= 1<<23); }
```

buf: ARRAY OF BITVECTOR(8)

SBVLT(buf[18], 0x00)

BVGE(buf[19]@buf[18]@buf[17]@buf[16], 0x00800000)

The Bad: Scalability Challenges



The Bad: Scalability Challenges

Application	Instrs/s	Queries/s	Solver %
[695	7.9	97.8
base64	20,520	42.2	97.0
chmod	5,360	12.6	97.2
comm	222,113	305.0	88.4
csplit	19,132	63.5	98.3
dircolors	1,019,795	4,251.7	98.6
echo	52	4.5	98.8
env	13,246	26.3	97.2
factor	12,119	22.6	99.7
join	1,033,022	3,401.2	98.1
ln	2,986	24.5	97.0
mkdir	3,895	7.2	96.6
Avg:	196,078	675.5	97.1

The Bad: Scalability Challenges

Application	Instrs/s	Queries/s	Solver %
[695	7.9	97.8
base64	20,520	42.2	97.0
chmod	5,360	12.6	97.2
comm	222,113	305.0	88.4
csplit	19,132	63.5	98.3
dircolors	1,019,795	4,251.7	98.6
echo	52	4.5	98.8
env	13,246	26.3	97.2
factor	12,119	22.6	99.7
join	1,033,022	3,401.2	98.1
ln	2,986	24.5	97.0
mkdir	3,895	7.2	96.6
Avg:	196,078	675.5	97.1

The Bad: Scalability Challenges

Eliminating irrelevant constraints

- In practice, each branch usually depends on a small number of variables

```
...
...
if (x < 10) {
    ...
}

```

PC: $x + y > 10 \wedge z \& -z = z$

$x < 10 ?$

→

The Bad: Scalability Challenges

Eliminating irrelevant constraints

- In practice, each branch usually depends on a small number of variables

```
...
...
PC:  $x + y > 10 \wedge$ 
 $x < 10 \wedge z < 10$ 
if (x < 10) {  $\longrightarrow$  x < 10?
    ...
}
```

The Bad: Scalability Challenges

Caching Solutions

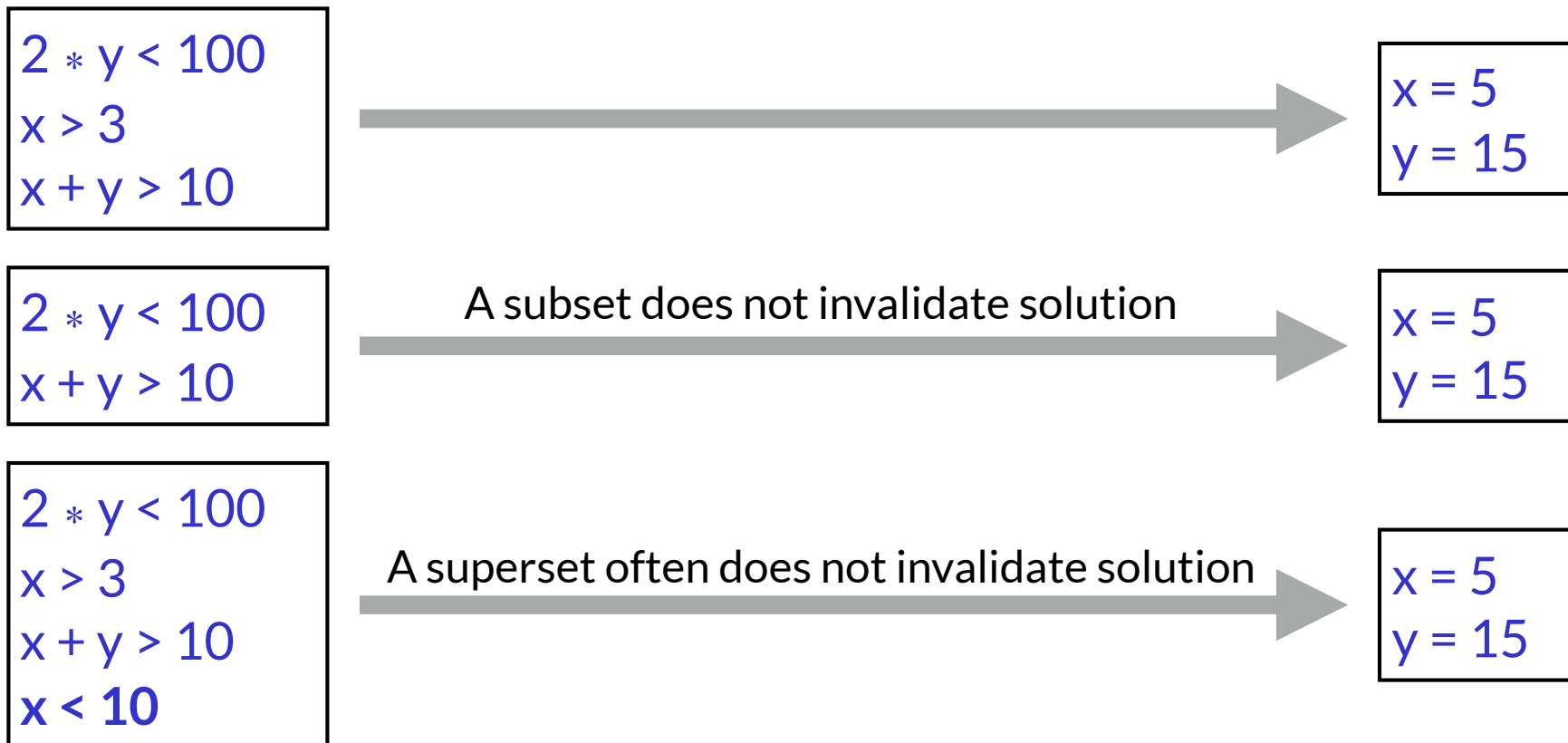
- Static set of branches: lots of similar constraint sets



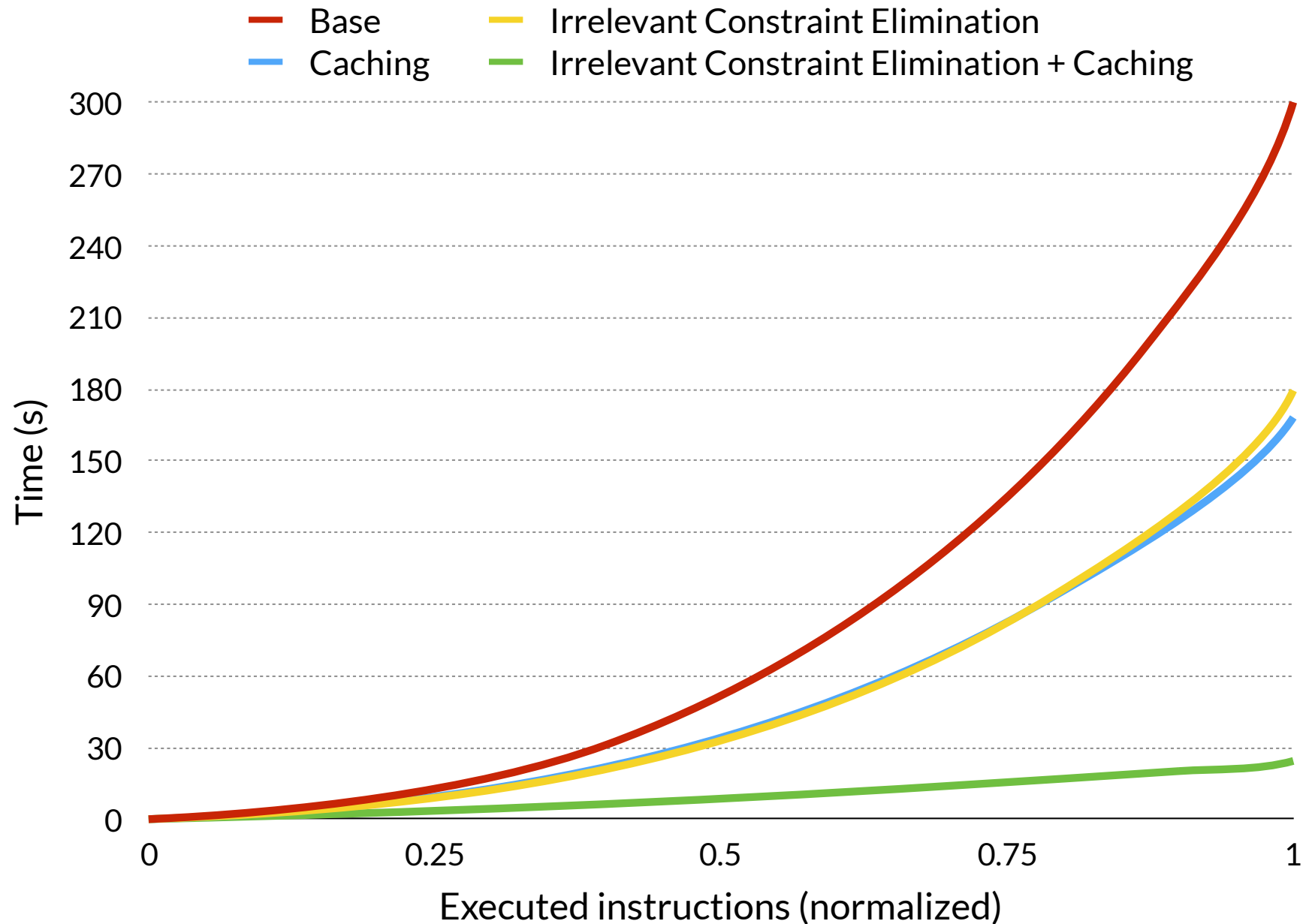
The Bad: Scalability Challenges

Caching Solutions

- Static set of branches: lots of similar constraint sets



The Bad: Scalability Challenges



The Ugly: Reasoning about Arrays



Joint work with Cristian Cadar, David Perry, Xiangyu Zhang

The Ugly: Reasoning about Arrays

Arrays are **pervasive** in software

- strings
- hash tables
- vectors
- pointer operations

Many programs generate **large** constraints involving arrays with **symbolic indexes**

The Ugly: Reasoning about Arrays

```
char equivClass[256] = {0, 1, 1, .. 10, 11, 12, 13 .. 1, 1, , 1};
char accept[298]     = {0, 0, .. 41, 32, 28, .. 23, 2, 2, 0};
unsigned check[580] = {0, 1, 1, .. 141, 142, 144, .. 297, 297};
unsigned base[302]  = {0, 0, .. 412, 422, .. 120, 395, 398};
unsigned def[302]   = {0, 297, .. 63, 300, .. 297, 297, 297};
char meta[54]       = {0, 1, 1, .. 1, 3, 3, .. 3, 3, 1, 1, 1};
unsigned next[580]  = {0, 6, 7, .. 132, 133, .. 297, 297};
```

```
Void tokenMatch(char *input) {
    unsigned currState = 0;
    char charPtr = input;
    do {
        char currClass = equivClass[*charPtr];
        if(accept[currState]) {
            lastAcceptState = currState;
            lastAcceptPos = charPtr;
        }
        while(check[base[currState] + currClass] != currState) {
            currState = def[currState];
            if(currState >= 298)
                currClass = meta[currClass];
        }
        currState = next[base[currState] + currClass];
        ++charPtr;
    } while(base[currState] != 526);
}
```


The Ugly: Reasoning about Arrays

```
char equivClass[256] = {0, 1, 1, .. 10, 11, 12, 13 .. 1, 1, , 1};
char accept[298]     = {0, 0, .. 41, 32, 28, .. 23, 2, 2, 0};
unsigned check[580] = {0, 1, 1, .. 141, 142, 144, .. 297, 297};
unsigned base[302]  = {0, 0, .. 412, 422, .. 120, 395, 398};
unsigned def[302]   = {0, 297, .. 63, 300, .. 297, 297, 297};
char meta[54]       = {0, 1, 1, .. 1, 3, 3, .. 3, 3, 1, 1, 1};
unsigned next[580]  = {0, 6, 7, .. 132, 133, .. 297, 297};
```

```
Void tokenMatch(char *input) {
    unsigned currState = 0;
    char charPtr = input;
    do {
        char currClass = equivClass[*charPtr];
        if(accept[currState]) {
            lastAcceptState = currState;
            lastAcceptPos = charPtr;
        }
        while(check[base[currState] + currClass] != currState) {
            currState = def[currState];
            if(currState >= 298)
                currClass = meta[currClass];
        }
        currState = next[base[currState] + currClass];
        ++charPtr;
    } while(base[currState] != 526);
}
```

KLEE 1h DFS mode → 20 paths

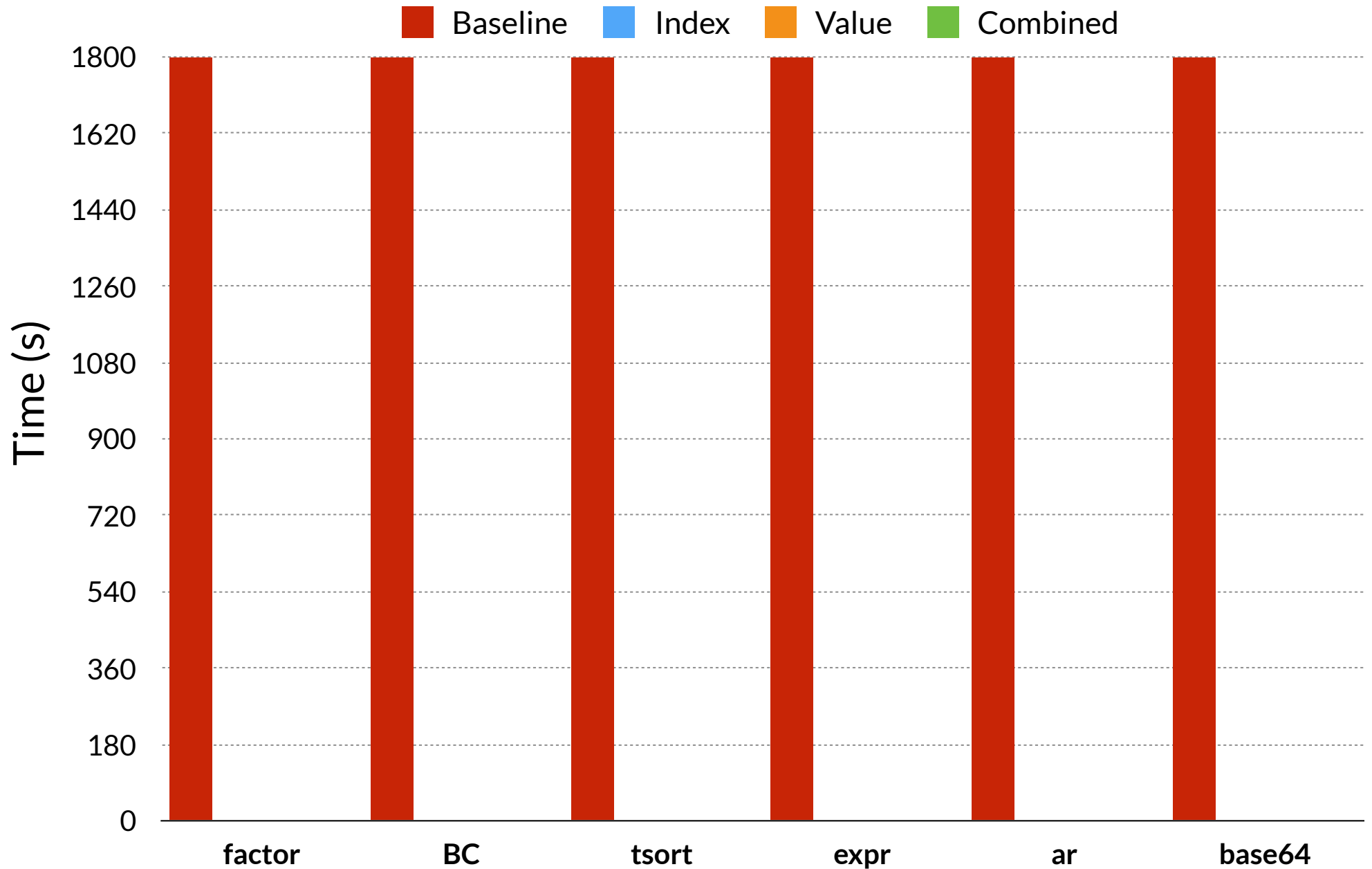
The Ugly: Reasoning about Arrays

```
char equivClass[256] = {0, 1, 1, .. 10, 11, 12, 13 .. 1, 1, , 1};
char accept[298]     = {0, 0, .. 41, 32, 28, .. 23, 2, 2, 0};
unsigned check[580] = {0, 1, 1, .. 141, 142, 144, .. 297, 297};
unsigned base[302]  = {0, 0, .. 412, 422, .. 120, 395, 398};
unsigned def[302]   = {0, 297, .. 63, 300, .. 297, 297, 297};
char meta[54]       = {0, 1, 1, .. 1, 3, 3, .. 3, 3, 1, 1, 1};
unsigned next[580]  = {0, 6, 7, .. 132, 133, .. 297, 297};
```

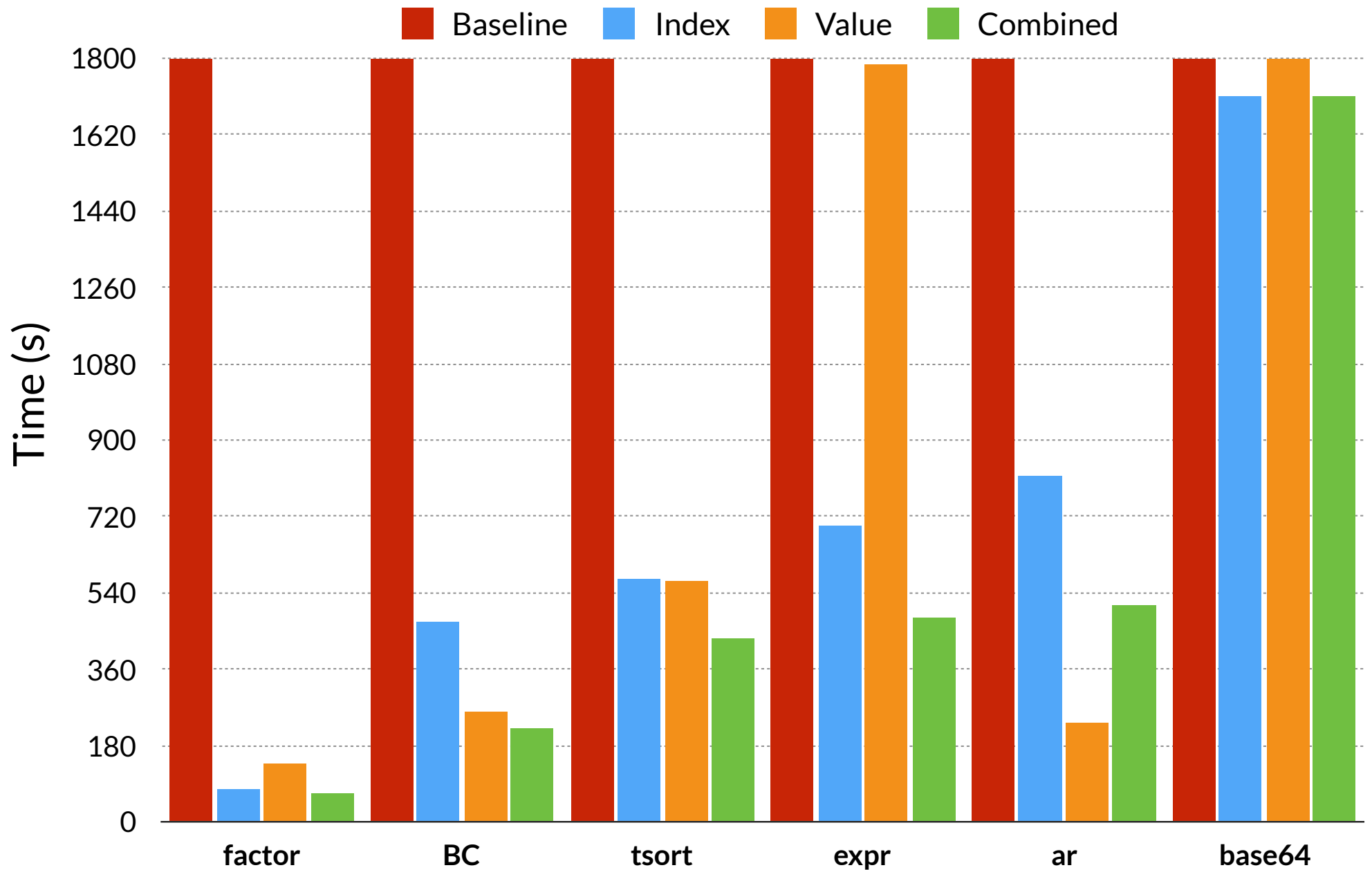
```
Void tokenMatch(char *input) {
    unsigned currState = 0;
    char charPtr = input;
    do {
        char currClass = equivClass[*charPtr];
        if(accept[currState]) {
            lastAcceptState = currState;
            lastAcceptPos = charPtr;
        }
        while(check[base[currState] + currClass] != currState) {
            currState = def[currState];
            if(currState >= 298)
                currClass = meta[currClass];
        }
        currState = next[base[currState] + currClass];
        ++charPtr;
    } while(base[currState] != 526);
}
```

KLEE 1h DFS mode → 828 paths

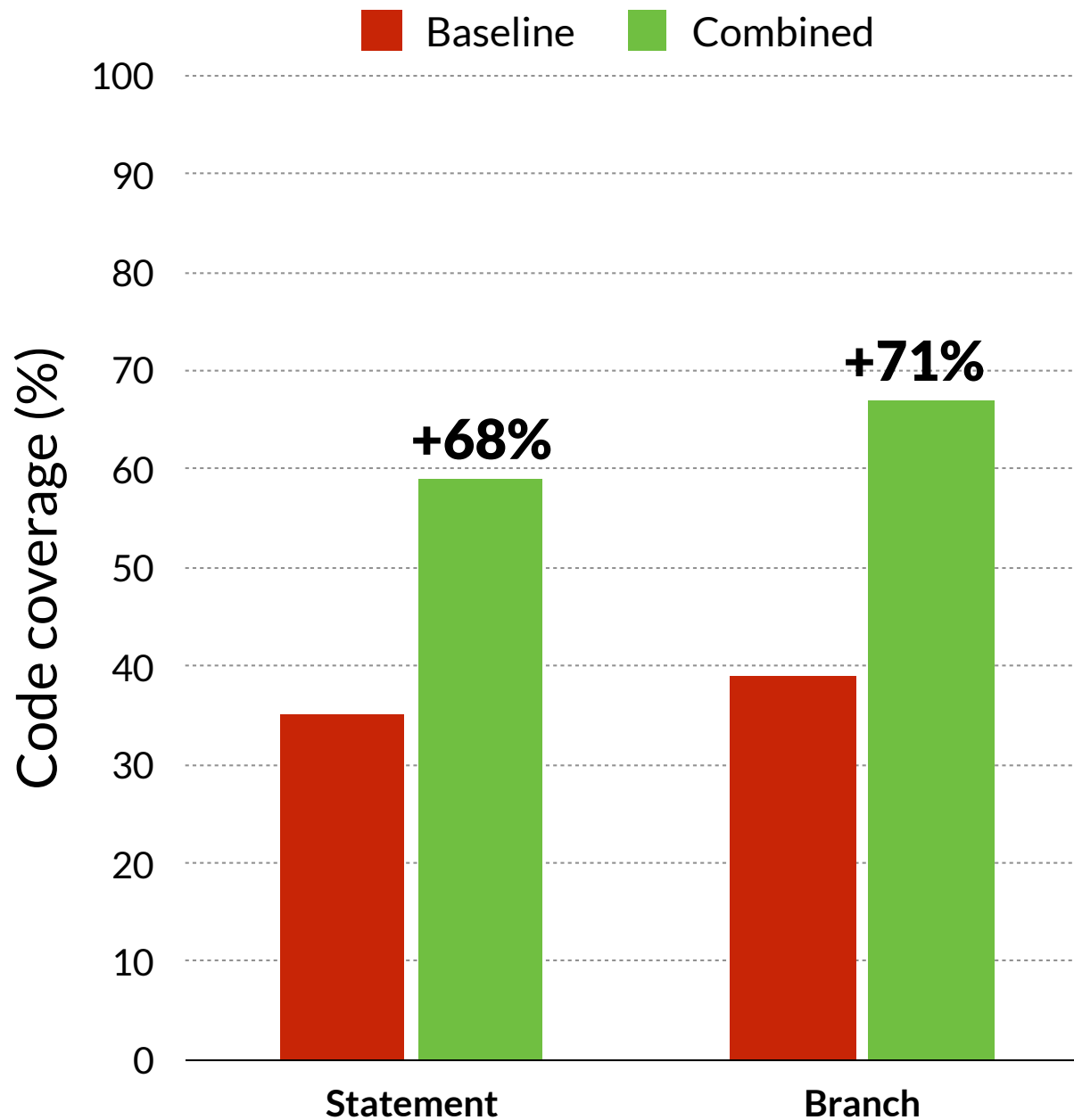
The Ugly: Reasoning about Arrays



The Ugly: Reasoning about Arrays



The Ugly: Reasoning about Arrays



The Good: Checks and Accuracy



The Bad: Scalability Challenges



The Ugly: Reasoning about Arrays



~~The Ugly~~: Reasoning about Arrays

