

# **WEAK INTERFACES → WEAK DEFENCES**

## **...THE BANE OF IMPLICIT CONVERSIONS IN OUR FUNCTION**

RICHÁRD SZALAY,  
ÁBEL SINKOVICS,  
ZOLTÁN PORKOLÁB

EÖTVÖS LORÁND UNIVERSITY, FACULTY OF INFORMATICS  
DEPARTMENT OF PROGRAMMING LANGUAGES & COMPILERS

**C++Now 2021, 2021. 05. 04.**

Prepared with the professional support of the Doctoral Student Scholarship Program of the  
Co-operative Doctoral Program of the Ministry of Innovation and Technology  
financed from the National Research, Development and Innovation Fund.





# THE ROUGH ROAD TOWARDS UPGRADING TO C++ MODULES

RICHÁRD SZALAY @Whisperity,  
ZOLTÁN PORKOLÁB

EÖTVÖS LORÁND UNIVERSITY  
FACULTY OF INFORMATICS  
DEPT. OF PROG. LANG. & COMPILERS

2019. 05. 10.

Research effort supported by the ÚNKP-18-2 "New National Excellence Program" of the Ministry of Human Capacities.  
Publication supported by the European Union, co-financed by the European Social Fund in project EFOP-3.6.3-VEKOP-16-2017-00002.



EMBERI ERŐFORRÁSOK  
MINISZTERIUM

SZÉCHENYI



## 1 Argument swapping

- “Name-based analysis”
- Reactiveness → proactiveness
- Existing guidelines

## 2 Implicit conversions

## 3 Type-based guards

- The Solution?
- The Analysis
- Empirica

## 4 Summary

# ARGUMENT SELECTION DEFECTS

Coined by Rice *et al.*<sup>1</sup>, but goes back longer.

Wrong argument (from available “set of expressions”) passed to function call.

```
int f(string host_name, int port, string message);  
  
string author = "Richard.Szalay";  
string greeting = "Hello, World!";  
  
f(author, 8080, greeting);
```

---

<sup>1</sup>Rice et al., “Detecting Argument Selection Defects”.

# ARGUMENT SWAPS ( $\subset$ ARGUMENT SELECTION DEFECTS)

Special case when arguments are as intended, but out of order.

Previous literature findings:

- adjacency increases chance of mistake
- too many parameters ( $\geq 5$ ) increase chance of mistake, too

```
int f2(string message, string hostName, int port);
```

# PROPER VARIABLE NAMES?

Name our variables properly!<sup>2345</sup>

---

<sup>2</sup>Liu et al., “Nomen est Omen: Exploring and Exploiting Similarities between Argument and Parameter Names”.

<sup>3</sup>Pradel and Gross, “Name-Based Analysis of Equally Typed Method Arguments”.

<sup>4</sup>Rice et al., “Detecting Argument Selection Defects”.

<sup>5</sup>Scott et al., “Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments”.

# PROPER VARIABLE NAMES?

Name our variables properly!<sup>2345</sup>

Simple ideas to compare:

```
void  $\mathcal{F}$ (int  $\mu$ , int  $\lambda$ ) { /* ... */ }  
//:  
foo( $\beta$ ,  $\vartheta$ );
```

---

<sup>2</sup>Liu et al., “Nomen est Omen: Exploring and Exploiting Similarities between Argument and Parameter Names”.

<sup>3</sup>Pradel and Gross, “Name-Based Analysis of Equally Typed Method Arguments”.

<sup>4</sup>Rice et al., “Detecting Argument Selection Defects”.

<sup>5</sup>Scott et al., “Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments”.

# “NAME-BASED” ANALYSIS

Simple ideas to compare:

```
void F(int μ, int λ) { /* ... */ }  
//:  
foo(β, ϑ);
```

- String equality, suffix/prefix coverage, pattern containment
- Edit distances (e.g. Levenshtein-distance, ...)
- Morpheme extraction<sup>6</sup>
- Typo analysis (e.g. key distance)

---

<sup>6</sup>Scott et al., “Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments”.

# TOOLS THAT EXIST

Link rot affects papers of previous slides... 😞

---

<sup>7</sup>Scott et al., “Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments”.

# TOOLS THAT EXIST

Link rot affects papers of previous slides... 😞

- GrammaTech, *SwapD*<sup>7</sup>.
- Clang-Tidy – hopefully upcoming: Varjú, Barancsuk, and Whisperity, *Add readability-suspicious-call-argument check*

---

<sup>7</sup>Scott et al., “Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments”.

```
diff --git a/gcc/config/i386/adxintrin.h b/gcc/config/i386/adxintrin.h
index 9c4152b9f360c0f9be408c84da4950ded8ad5654..7acdaf4ab6f7675401eab1c512175e6493620
--- a/gcc/config/i386/adxintrin.h
+++ b/gcc/config/i386/adxintrin.h
@@ -33,7 +33,7 @@ __attribute__((__gnu_inline__, __always_inline__, __artificial__))
_subborrow_u32 (unsigned char __CF, unsigned int __X,
                 unsigned int __Y, unsigned int *__P)
{
-  return __builtin_ia32_sbb_u32 (__CF, __Y, __X, __P);
+  return __builtin_ia32_sbb_u32 (__CF, __X, __Y, __P);
}

extern __inline unsigned char
@@ -58,7 +58,7 @@ __attribute__((__gnu_inline__, __always_inline__, __artificial__))
_subborrow_u64 (unsigned char __CF, unsigned long long __X,
                  unsigned long long __Y, unsigned long long *__P)
{
-  return __builtin_ia32_sbb_u64 (__CF, __Y, __X, __P);
+  return __builtin_ia32_sbb_u64 (__CF, __X, __Y, __P);
}

extern __inline unsigned char
```

GNU Compiler Collection

---

andreser, *\_subborrow\_u64 argument order inconsistent*



# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?
  - ▶ `fun(*objw, objw→getSome());`

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?
  - ▶ `fun(*objw, objw→getSome());`
  - ▶ `fun(1, 2);`

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?
  - ▶ `fun(*objw, objw→getSome());`
  - ▶ `fun(1, 2);`
  - ▶ `fun(variance, 4 + 8);`

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?
  - ▶ `fun(*objw, objw→getSome());`
  - ▶ `fun(1, 2);`
  - ▶ `fun(variance, 4 + 8);`
- Oh and...

# THE DOWNSIDE OF NAMES

- You need to “have them” (or figure them out)
- Complex/unnamed expressions?
  - ▶ `fun(*objw, objw→getSome());`
  - ▶ `fun(1, 2);`
  - ▶ `fun(variance, 4 + 8);`
- Oh and...?    **void** someAPIFunction(**int**, **int**, **long**);

src/backend/executor/execMain.c

```
722 ExecCheckRTEPermsModified(Oid relOid, Oid userid, Bitmapset *modifiedCols,
723                             AclMode requiredPerms)
724 {
725     int          col = -1;
726
727     /*
728      * When the query doesn't explicitly update any columns, allow the query
729      * if we have permission on any column of the rel. This is to handle
730      * SELECT FOR UPDATE as well as possible corner cases in UPDATE.
731      */
732     if (bms_is_empty(modifiedCols))
733     {
734         if (pg_attribute_aclcheck_all(relOid, userid, requiredPerms,
735                                         ACLMASK_ANY) != ACLCHECK_OK)
736             return false;
737     }
    
```

2 < 1st argument 'relOid' (passed to 'table\_oid') looks like it might be swapped with the 2nd, 'userid' (passed to 'roleid')

src/include/utils/acl.h

```
260                                     Oid roleid, AclMode mode);  
261 extern AclResult pg_attribute_aclcheck_all(Oid table_oid, Oid roleid,  
262                                             AclMode mode, AclMaskHow how);  
263 extern AclResult pg_class_aclcheck(Oid table_oid, Oid roleid, AclMode mode);  
264 extern AclResult pg_database_aclcheck(Oid db_oid, Oid roleid, AclMode mode);  
265 extern AclResult pg_proc_aclcheck(Oid proc_oid, Oid roleid, AclMode mode);  
266 extern AclResult pg_language_aclcheck(Oid lang_oid, Oid roleid, AclMode mode);
```

1

in the call to 'pg\_attribute\_aclcheck\_all', declared here >

- **reloid** – relation (table) **owner**(?) **ID**
- **roleid** – **role** (user) **ID**

# REACTIVE → PROACTIVE

$g(T, V);$



void  $g(\text{int velocity, int torque})$

# REACTIVE → PROACTIVE

$g(T, V);$



void  $g(\text{int velocity, int torque})$

- Detect call sites...after-the-fact

# REACTIVE → PROACTIVE

$g(T, V);$



void  $g(\text{int velocity, int torque})$

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report

# REACTIVE → PROACTIVE

$g(T, V);$



```
void g(int velocity, int torque)
```

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report
- False positives?

# REACTIVE → PROACTIVE

```
g(T, V);  
↓      ↓  
void g(int velocity, int torque)
```

```
void g2(velocity_t V,  
          torque_t T);
```

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report
- False positives?

# REACTIVE → PROACTIVE

```
g(T, V);  
      ↓   ↓  
void g(int velocity, int torque)
```

```
void g2(velocity_t V,  
        torque_t T);
```

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report
- False positives?

- Use the type system to guard us!

# REACTIVE → PROACTIVE

```
g(T, V);  
      ↓   ↓  
void g(int velocity, int torque)
```

```
void g2(velocity_t V,  
          torque_t T);
```

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report
- False positives?

- Use the type system to guard us!
- Swapped expressions → *compile error*

# REACTIVE → PROACTIVE

```
g(T, V);  
↓      ↓  
void g(int velocity, int torque)
```

```
void g2(velocity_t V,  
          torque_t T);
```

- Detect call sites...after-the-fact
- Significant enough mismatch **in name** → report
- False positives?

- Use the type system to guard us!
- Swapped expressions → *compile error*
- Pure developer intent!

## diff.c

```
3347 static void emit_binary_diff(struct diff_options *o,
3348     mmfile_t *one, mmfile_t *two)
3349 {
3350     emit_diff_symbol(o, DIFF_SYMBOL_BINARY_DIFF_HEADER, NULL, 0, 0);
3351     emit_binary_diff_body(o, one, two);
3352     emit_binary_diff_body(o, two, one);
3353 }
```

2 < 2nd argument 'two' (passed to 'one') looks like it might be swapped with the  
3rd, 'one' (passed to 'two')

## I.24: Avoid adjacent unrelated parameters of the same type

### Reason

Adjacent arguments of the same type are easily swapped by mistake.

### Example, bad

Consider:

```
void copy_n(T* p, T* q, int n); // copy from [p:p + n] to [q:q + n]
```

This is a nasty variant of a K&R C-style interface. It is easy to reverse the "to" and "from" arguments.

Use `const` for the "from" argument:

```
void copy_n(const T* p, T* q, int n); // copy from [p:p + n] to [q:q + n]
```

### Exception

If the order of the parameters is not important, there is no problem:

```
int max(int a, int b);
```

### Alternative

Don't pass arrays as pointers, pass an object representing a range (e.g., a `span`):

```
void copy_n(span<const T> p, span<T> q); // copy from p to q
```

### Alternative

Define a `struct` as the parameter type and name the fields for those parameters accordingly:

```
struct SystemParams {  
    string config_file;  
    string output_path;  
    seconds timeout;  
};  
void initialize(SystemParams p);
```

This tends to make invocations of this clear to future readers, as the parameters are often filled in by name at the call site.

### Enforcement

(Simple) Warn if two consecutive parameters share the same type.

## I.24: Avoid adjacent parameters that can be invoked by the same arguments in either order with different meaning

### Alternative

Define a `struct` as the parameter type and name the fields for those parameters accordingly:

```
struct SystemParams {  
    string config_file;  
    string output_path;  
    seconds timeout;  
};  
void initialize(SystemParams p);
```

### Reason

Adjacent arguments of the same type are easily swapped by mistake.

### Example, bad

Consider:

```
void copy_n(T* p, T* q, int n); // copy from [p:p + n) to [q:q + n)
```

This is a nasty variant of a K&R C-style interface. It is easy to reverse the "to" and "from" arguments.

Use `const` for the "from" argument:

```
void copy_n(const T* p, T* q, int n); // copy from [p:p + n) to [q:q + n)
```

### Exception

If the order of the parameters is not important, there is no problem:

```
int max(int a, int b);
```

This tends to make invocations of this clear to future readers, as the parameters are often filled in by name at the call site.

### Note

Only the interface's designer can adequately address the source of violations of this guideline.

### Enforcement strategy

(Simple) Warn if two consecutive parameters share the same type

We are still looking for a less-simple enforcement.

### Alternative

Don't pass arrays as pointers, pass an object representing a range (e.g., a `span`):

```
void copy_n(span<const T> p, span<T> q); // copy from p to q
```

# **TYPES!**

C++ IS UNLIKE JAVA<sup>8</sup>

# TYPES!

- **typedef/using** (?)

---

<sup>8</sup>Most of the existing splutions are for Java

## TYPES!

- **typedef/using** ?
- **const** T& troubles!

---

<sup>8</sup>Most of the existing splutions are for Java

# TYPES!

- **typedef/using** (?)
- **const T&** troubles!
- **const T ≡ T** (?)

---

<sup>8</sup>Most of the existing splutions are for Java

# TYPES!

- **typedef/using** (?)
- **const** T& troubles!
- **const** T ≡ T (?)
- Implicit conversions...

---

<sup>8</sup>Most of the existing splutions are for Java

# WILL THIS BE CAUGHT?

```
struct Complex { double Re, Im; /* ... */ };
void h(int Scalar, Complex Comp);

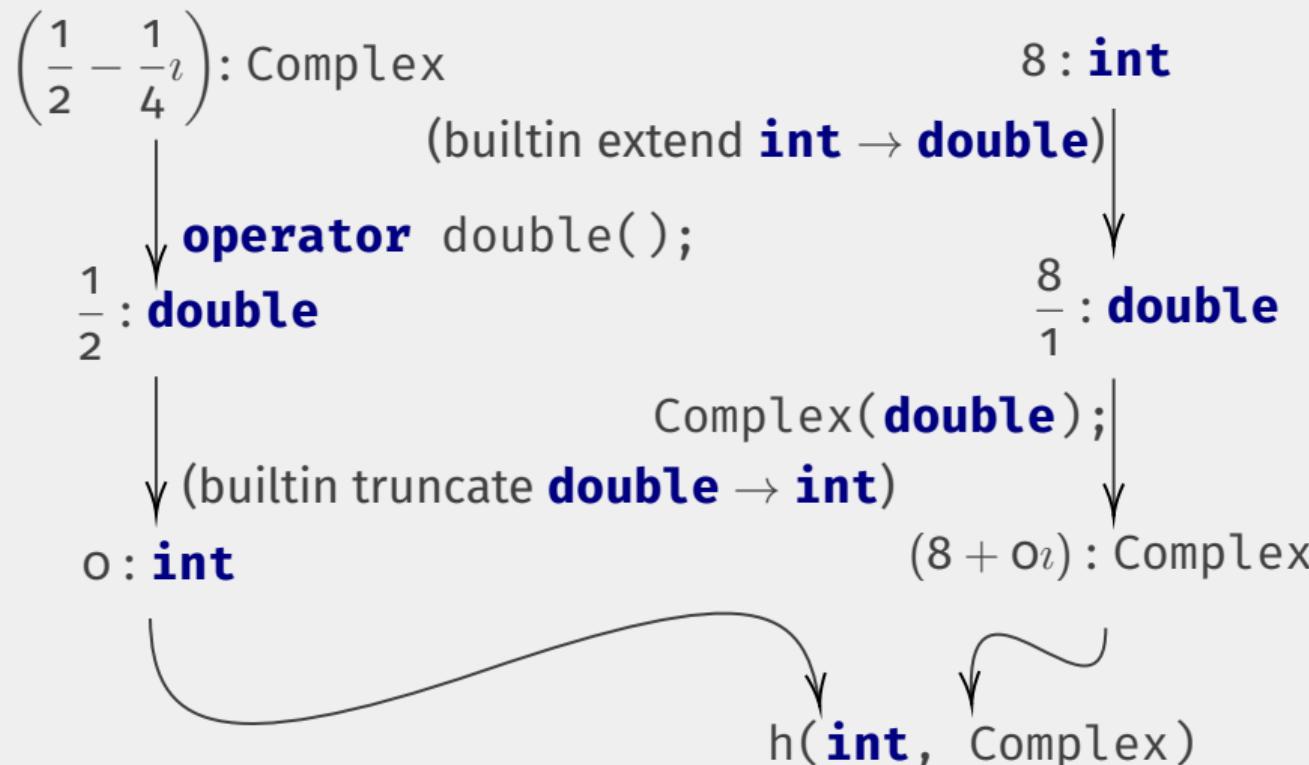
void test() {
    int S = 8;
    Complex C = Complex{.5f, -.25f}; // = ( $\frac{1}{2} - \frac{1}{4}i$ )
    h(C, S); // ← ?
}
```

# IT DEPENDS ...

```
struct Complex {
    double R, I;
    Complex(double real) : R(real), I(0.0) {}
    operator double() const { return R; }
    // :
};

void h(int Scalar, Complex Comp);
void test() {
    int S = 8;
    Complex C = Complex{.5f, -.25f}; // = ( $\frac{1}{2} - \frac{1}{4}i$ )
    h(C, S); // ✓
}
```

# IMPLICIT CONVERSIONS



# SUBTLE MISTAKE, EXPANDED

```
struct Complex {
    double R, I;
    Complex(double real) : R(real), I(0.0) {}
    operator double() const { return R; }
    // :
};

void h(int Scalar, Complex Comp);

void test() {
    h(Complex{.5f, -.25f}, 8); // ≡ h(C, S); from before...
    h(0, Complex{8, 0});
}
```

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

1. **Maybe standard conversion sequence:**

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

1. **Maybe standard conversion sequence:**

- 1.1 *Maybe decay* (lvalue→rvalue, array/function→pointer)
- 1.2 *Maybe numeric adjustment*

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. **Maybe standard conversion sequence:**

- 1.1 *Maybe decay* (lvalue → rvalue, array/function → pointer)
- 1.2 *Maybe numeric adjustment*
  - integral ↔ integral, enum → integral

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. **Maybe standard conversion sequence:**

1.1 *Maybe decay* (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 *Maybe numeric adjustment*

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. **Maybe standard conversion sequence:**

1.1 *Maybe decay* (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 *Maybe numeric adjustment*

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class, T\*  $\rightarrow$  **void\***

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. **Maybe standard conversion sequence:**

1.1 *Maybe decay* (lvalue → rvalue, array/function → pointer)

1.2 *Maybe numeric adjustment*

- integral ↔ integral, enum → integral
- floating ↔ integral
- derived → base class, T\* → **void\***
- **null**-constants → pointer (?!)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue→rvalue, array/function→pointer)

1.2 Maybe numeric adjustment

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void\***
- **null**-constants  $\rightarrow$  pointer (?!)
- Anything you can imagine  $\rightarrow$  **bool** (!!)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue→rvalue, array/function→pointer)

1.2 Maybe numeric adjustment

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void\***
- **null**-constants  $\rightarrow$  pointer (?!)
- Anything you can imagine  $\rightarrow$  **bool** (!!)

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue→rvalue, array/function→pointer)

1.2 Maybe numeric adjustment

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void\***
- **null**-constants  $\rightarrow$  pointer (?!)
- Anything you can imagine  $\rightarrow$  **bool** (!!)

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence*  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  exists and defined as:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue→rvalue, array/function→pointer)

1.2 Maybe numeric adjustment

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void\***
- **null**-constants  $\rightarrow$  pointer (?!)
- Anything you can imagine  $\rightarrow$  **bool** (!!)

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

## 2. Maybe user-defined conversion (one function call!)

## 3. Maybe standard conversion sequence (same as above)

...if the path taken *uniquely exists*.



# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. *Maybe standard conversion sequence:*

1.1 *Maybe* decay (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 *Maybe* numeric adjustment

- integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral
- floating  $\leftrightarrow$  integral
- derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void** $^*$
- **null**-constants  $\rightarrow$  pointer
- Anything you can imagine  $\rightarrow$  **bool**

1.3 *Maybe* function pointer adjustment (“lose **noexcept**”)

1.4 *Maybe* qualifier adjustment (“gain **const volatile**”)

2. *Maybe* user-defined conversion (one function call!)

3. *Maybe* standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 Maybe numeric adjustment

- ~~integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral~~
- ~~floating  $\leftrightarrow$  integral~~
- derived  $\rightarrow$  base class,  $T^* \rightarrow \mathbf{void}^*$
- **null**-constants  $\rightarrow$  pointer
- Anything you can imagine  $\rightarrow \mathbf{bool}$

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 Maybe numeric adjustment

- ~~integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral~~
- ~~floating  $\leftrightarrow$  integral~~
- derived  $\rightarrow$  base class, ~~T\*~~  $\rightarrow$  **void\***
- **null**-constants  $\rightarrow$  pointer
- Anything you can imagine  $\rightarrow$  **bool**

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue $\rightarrow$ rvalue, array/function $\rightarrow$ pointer)

1.2 Maybe numeric adjustment

- ~~integral  $\leftrightarrow$  integral, enum  $\rightarrow$  integral~~
- ~~floating  $\leftrightarrow$  integral~~
- ~~derived  $\rightarrow$  base class,  $T^*$   $\rightarrow$  **void**\*~~
- **null**-constants  $\rightarrow$  pointer
- ~~Anything you can imagine  $\rightarrow$  bool~~

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (lvalue → rvalue, ~~array~~/function → pointer)

1.2 Maybe numeric adjustment

- ~~integral ↔ integral, enum → integral~~
- ~~floating ↔ integral~~
- ~~derived → base class, T\* → void\*~~
- **null**-constants → pointer
- ~~Anything you can imagine → bool~~

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

# IMPLICIT CONVERSIONS

An *implicit conversion sequence* is a unique path from  $\mathcal{T}_1 \rightarrow \mathcal{T}_2$  s.t.:

## 1. Maybe standard conversion sequence:

1.1 Maybe decay (~~lvalue → rvalue~~, ~~array/function~~ → pointer)

1.2 Maybe numeric adjustment

- ~~integral ↔ integral, enum → integral~~
- ~~floating ↔ integral~~
- ~~derived → base class, T\* → void\*~~
- ~~null-constants → pointer~~
- ~~Anything you can imagine → bool~~

1.3 Maybe function pointer adjustment (“lose **noexcept**”)

1.4 Maybe qualifier adjustment (“gain **const volatile**”)

2. Maybe user-defined conversion (one function call!)

3. Maybe standard conversion sequence (same as above)

**Types?**

**Strong types!**

**Stronger types!**

**Strong...err...types?**

# THE SOLUTION



## Tuesday, May 4

7:00am MDT

Converting a State Machine to a C++ 20 Coroutine

Steve Downey

Range-Based Text Formatting - For a Future Range-Based Standard Library

Eric Sander

Code Analysis++

Intermediate Knowledge

9:00am MDT

C++ Standard Parallelism

Bryce Adelman-Lieback

Interactive C++ in a Jupyter Notebook Using Modules for Incremental Compilation

Steven Brandt

Simplest Safe Integers

Peter Sommerlad

## Thursday, May 6

7:00am MDT

So You Think You Know How to Work With Concepts?

Andrea Krammer

CMake + Conan: 3 Years Later

Maxim Parf

Simplest Strong Typing instead of Language Proposal ( P0109 )

Peter Sommerlad

9:00am MDT

Algorithms from a Compiler Developer's Toolbox

Gábor Horváth

Windows, MacOS and the Web: Lessons from Cross-platform Development at think-cell

Detlev Töpfer

Variations on variants

Rui Barros

11:30am MDT

Designing Concurrent C++ Applications

Lukas Reiter

C++ Insights: How Stuff Works, Lambdas and More!

Andreas Ferstl

Library Approaches for Strong Type Aliases

Anthony Williams

# MIXABLE ADJACENT PARAMETERS

```
void  
p (int i, int j, double d, Complex c, std::string s);
```

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  *Same type, trivially*

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  *Same type, trivially*
- Is **int** mixable with **double**?

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  *Same type, trivially*
- Is **int** mixable with **double**?  *standard conversion*

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?  **standard + user**

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?  **standard + user**
- Is **double** mixable with Complex?

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?  **standard + user**
- Is **double** mixable with Complex?  **user conversion**

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?  **standard + user**
- Is **double** mixable with Complex?  **user conversion**
- Is Complex mixable with std::string?

# MIXABLE ADJACENT PARAMETERS

**void**

```
p (int i, int j, double d, Complex c, std::string s);
```

- Is **int** mixable with **int**?  **Same type, trivially**
- Is **int** mixable with **double**?  **standard conversion**
- Is **int** mixable with Complex?  **standard + user**
- Is **double** mixable with Complex?  **user conversion**
- Is Complex mixable with std::string?  **No**

# MIXABLE ADJACENT PARAMETERS

```
void  
p (int i, int j, double d, Complex c, std::string s);  
^~~~~~^~~~~~^~~~~~^~~~~~^
```

- Is **int** mixable with **int**? **Same type, trivially**
- Is **int** mixable with **double**? **standard conversion**
- Is **int** mixable with Complex? **standard + user**
- Is **double** mixable with Complex? **user conversion**
- Is Complex mixable with std::string? **No**
- ... (at most  $\mathcal{O}\left(\frac{n(n-1)}{2}\right)$  checks)

# ANALYSIS IN CLANG-TIDY

Implemented automated analysis rule in Clang-Tidy.<sup>9</sup>  
In the following we'll discuss 7 C and 9 C++ projects:

- curl
- git
- netdata
- PHP
- PostgreSQL
- Redis
- tmux
- Bitcoin
- CodeCompass
- guetzli
- LLVM
- OpenCV
- Protobuf
- Tesseract
- Xerces
- Z3

---

<sup>9</sup>Whisperity, Add bugprone-easily-swappable-parameters check.

# ANALYSIS IN CLANG-TIDY

Implemented automated analysis rule in Clang-Tidy.<sup>9</sup>  
In the following we'll discuss 7 C?? and 9 C++ projects:

- curl
- git
- netdata
- PHP
- PostgreSQL
- Redis
- tmux
- Bitcoin
- CodeCompass
- guetzli
- LLVM
- OpenCV
- Protobuf
- Tesseract
- Xerces
- Z3

---

<sup>9</sup>Whisperity, Add bugprone-easily-swappable-parameters check.

Name	Analyzer statistics	Latest storage date	Analysis duration ↓
lvm-project_len2-cvr-imp-rel-fil	clang-tidy: ✓ (2704)	2021-01-19 11:29:03	00:50:53
openpy_len2-cvr-imp-rel-fil	clang-tidy: ✓ (1405)	2021-01-19 19:30:47	00:08:30
bitcoin_len2-cvr-imp-rel-fil	clang-tidy: ✓ (483)	2021-01-18 16:00:18	00:04:43
z3_len2-cvr-imp-rel-fil	clang-tidy: ✓ (735)	2021-01-19 21:39:07	00:04:02
protobuf_len2-cvr-imp-rel-fil	clang-tidy: ✓ (321)	2021-01-18 17:15:59	00:01:57
tesseract_len2-cvr-imp-rel-fil	clang-tidy: ✓ (475) ✘ (8)	2021-02-07 12:21:27	00:01:45
CodeCompass_len2-cvr-imp-rel-fil	clang-tidy: ✓ (107)	2021-01-19 09:21:06	00:01:15
php-src_len2-cvr-imp-rel-fil	clang-tidy: ✓ (756)	2021-01-18 16:59:11	00:00:56
postgres_len2-cvr-imp-rel-fil	clang-tidy: ✓ (1081)	2021-01-19 20:58:14	00:00:41
git_len2-cvr-imp-rel-fil	clang-tidy: ✓ (419)	2021-01-18 16:47:28	00:00:16
netdata_len2-cvr-imp-rel-fil	clang-tidy: ✓ (131) ✘ (109)	2021-02-07 12:16:42	00:00:08
curl_len2-cvr-imp-rel-fil	clang-tidy: ✓ (324)	2021-01-18 16:44:01	00:00:06
tmux_len2-cvr-imp-rel-fil	clang-tidy: ✓ (136)	2021-01-18 17:36:52	00:00:04
redis_len2-cvr-imp-rel-fil	clang-tidy: ✓ (116)	2021-01-18 17:35:16	00:00:02
quetzli_len2-cvr-imp-rel-fil	clang-tidy: ✓ (21)	2021-01-18 16:52:52	00:00:01

```
1738 * case of error.  
1739 */  
1740 int Classify::MakeNewTemporaryConfig(ADAPT_TEMPLATES Templates,  
1741                                     CLASS_ID ClassId,  
1742                                     int FontinfoId,  
1743                                     int NumFeatures,  
1744                                     INT_FEATURE_ARRAY Features,  
1745                                     FEATURE_SET FloatFeatures) {  
  
    1 < the first parameter in this range is 'ClassId' >  
    3 < after resolving type aliases, type of parameter 'ClassId' is 'int' >  
    6 < 3 adjacent parameters for 'MakeNewTemporaryConfig' of similar type are easily  
       swapped by mistake  
    4 < after resolving type aliases, type of parameter 'FontinfoId' is 'int' >  
    2 < the last parameter in this range is 'NumFeatures' >  
    5 < after resolving type aliases, type of parameter 'NumFeatures' is 'int' >
```

**Figure:** One finding from Tesseract<sup>10</sup> visualised using CodeChecker<sup>11</sup>.

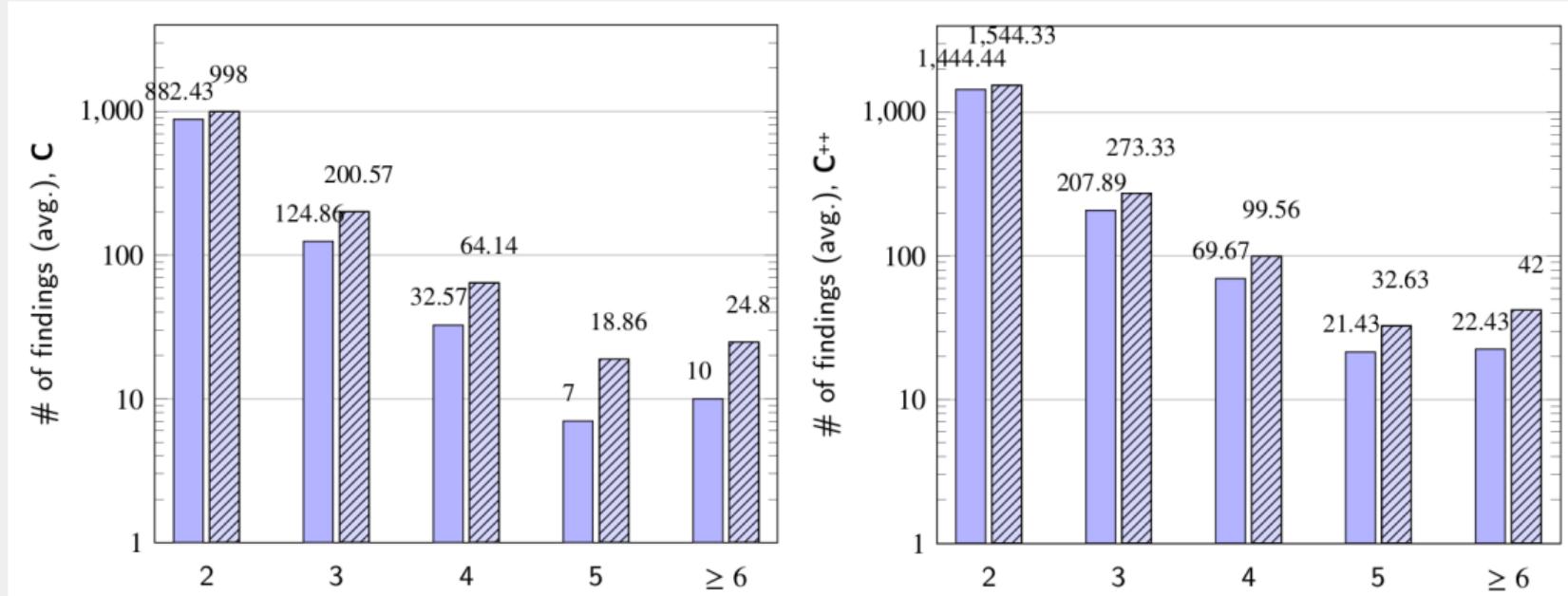
<sup>10</sup>Smith, Google, Inc., et al., *Tesseract OCR Engine*.

<sup>11</sup>Ericsson AB, *CodeChecker – an analyzer tooling, defect database and viewer*.

modules/calib3d/src/calibration.cpp

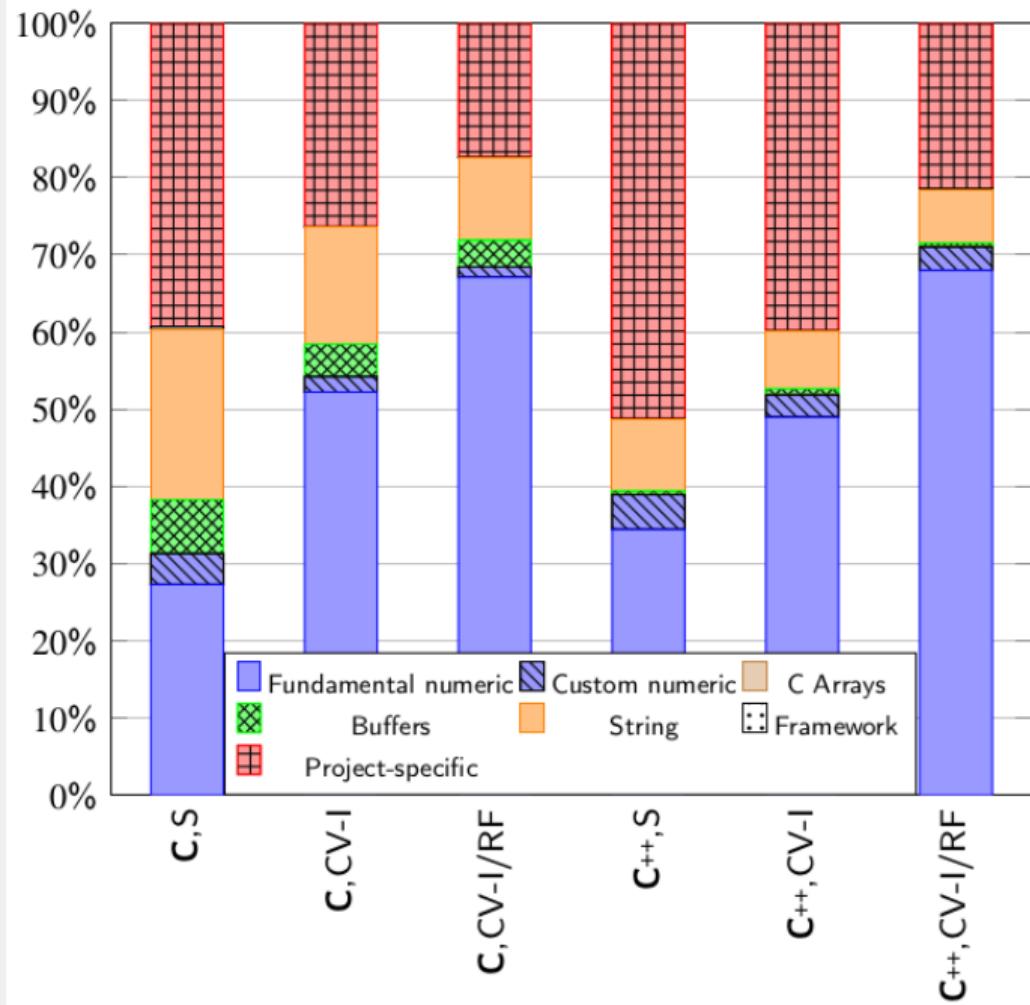
```
4167 float cv::rectify3Collinear( InputArray cameraMatrix1, InputArray _distCoeffs1,
1      < 1 the first parameter in this range is '_cameraMatrix1' >
2
3      < 8 adjacent parameters for 'rectify3Collinear' of similar type are easily swapped
4      by mistake
5
4168 InputArray _cameraMatrix2, InputArray _distCoeffs2,
4169 InputArray _cameraMatrix3, InputArray _distCoeffs3,
4170 InputArrayOfArrays imgpt1,
6      < 3 at a call site, 'const _InputArray &' might bind with same force as
7      'cv::InputArray'
8
4171 InputArrayOfArrays imgpt3,
9      < 2 2 the last parameter in this range is '_imgpt3' >
10
11      < 4 at a call site, 'const _InputArray &' might bind with same force as
12      'cv::InputArray'
13
4172 Size imageSize, InputArray _Rmat12, InputArray _Tmat12,
4173 InputArray _Rmat13, InputArray _Tmat13,
4174 OutputArray _Rmat1, OutputArray _Rmat2, OutputArray _Rmat3,
4175 OutputArray _Pmat1, OutputArray _Pmat2, OutputArray _Pmat3,
4176 OutputArray _Qmat,
4177 double alpha, Size newImgSize,
4178 Rect* roi1, Rect* roi2, int flags )
4179 {
10 // first, rectify the 1-2 stereo pair
```

# MORE PARAMETERS → MORE CHANCE OF MISTAKE



**Figure:** Avg. # of functions with problem. █ Strict mode, █ Generous **const** and implicit conversions.

Lang.	Project	Functions analysed (A)	Strict (S)			CV			Imp			CV ∪ Imp		
			T (total)	% of A		T	% of A	vs. S	T	% of A	vs. S	T	% of A	vs. CV
C	curl	865	134	15.49%	153	17.69%	19	209	24.16%	75	228	26.36%	75	19
	git	5 641	1 418	25.14%	1 466	25.99%	48	1 595	28.28%	177	1 644	29.14%	178	49
	netdata	719	227	31.57%	243	33.80%	16	294	40.89%	67	308	42.84%	65	14
	PHP	5 984	1 272	21.26%	1 304	21.79%	32	1 509	25.22%	237	1 539	25.72%	235	30
	Postgres	9 436	2 696	28.57%	2 804	29.72%	108	3 708	39.30%	1 012	3 820	40.48%	1 016	112
	Redis	1 745	393	22.52%	418	23.95%	25	454	26.02%	61	484	27.74%	66	36
	TMux	1 032	248	24.03%	259	25.10%	11	298	28.88%	50	306	29.65%	47	8
C++	Bitcoin	1 773	394	22.22%	412	23.24%	18	499	28.14%	105	512	28.88%	109	22
	CodeCompass	191	27	14.14%	27	14.14%	0	28	14.66%	1	28	14.66%	1	0
	guetzli	153	72	47.06%	76	49.67%	4	75	49.02%	3	81	52.94%	5	6
	LLVM-CTE	32 339	6 109	18.89%	6 181	19.11%	72	6 817	21.08%	704	6 898	21.33%	713	81
	OpenCV	6 746	3 286	48.71%	3 414	50.61%	128	3 590	53.22%	304	3 714	55.05%	300	124
	ProtoBuf	1 997	313	15.67%	317	15.87%	4	386	19.33%	73	392	19.63%	75	6
	Tesseract	1 962	793	40.42%	797	40.62%	4	874	44.55%	81	880	44.85%	83	6
	Xerces	1 594	446	27.98%	462	28.98%	16	465	29.17%	19	532	33.38%	70	67
	Z3	9 673	2 600	26.88%	2 608	26.96%	8	2 793	28.87%	193	2 801	28.96%	193	8



Whisperity!

# TOO MUCH REPORTS

- **int** max(**int** left, **int** right)
- **string** msgEmit(**string** txt1, **string** txt2, **string** txt3)

**“AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE”**

## **“AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE”**

*... can be invoked ... either order ... different meaning ...*

## **“AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE”**

*... can be invoked ... either order ... different meaning ...*

```
if (a < b) {  
    x + y;  
}
```

## **"AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE"**

*... can be invoked ... either order ... different meaning ...*

```
if (a < b) {          foo(1, a);
    x + y;           foo(2, b);
}
```

## "AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE"

*... can be invoked ... either order ... different meaning ...*

```
if (a < b) {          foo(1, a);
    x + y;           foo(2, b);
}
```

```
if (!xyz)
    return Default;
return *xyz;
```

# "AVOID ADJACENT UNRELATED PARAMETERS OF SAME TYPE"

... can be invoked ... either order ... different meaning ...

```
if (a < b) {  
    x + y;  
}
```

```
foo(1, a);  
foo(2, b);
```

```
if (!xyz)  
    return Default;  
return *xyz;
```

```
std::string LName = LHS→Name;  
std::string RName = RHS→Name;  
if (LName ≥ RName) { ...
```

# MAKE USE OF NAMES...

```
string concat(string str1, string str2)
```

② *What if we somehow realised that the order  
doesn't matter?...*

# MAKE USE OF NAMES...

```
string concat(string str1, string str2)
error format(info errorA, info errorB)
ordering compare(string LHS, string RHS)
matrix rectify(matrix Qmat, matrix Rmat, matrix Tmat)
```

Lang.	Project	Strict (S)		CV $\cup$ Imp		No-bool		Rel				Fil				Rel $\cap$ Fil			
		Total	Total	- vs S.	-	-	-	- vs. S	- vs. CV $\cup$ Imp	-	- vs. S	-	- vs. CV $\cup$ Imp	-	- vs. S	-	- vs. CV $\cup$ Imp		
C	curl	134	228	1	0.75%	61	45.52%	80	35.09%	4	2.99%	4	1.75%	64	47.76%	83	36.40%		
	git	1 418	1 644	0		826	58.25%	887	53.95%	142	10.01%	119	7.24%	902	63.61%	953	57.97%		
	netdata	227	308	0		123	54.19%	134	43.51%	6	2.64%	5	1.62%	124	54.63%	135	43.83%		
	PHP	1 272	1 539	6	0.47%	628	49.37%	699	45.42%	165	12.97%	164	10.66%	697	54.80%	766	49.77%		
	Postgres	2 696	3 820	157	5.82%	1 449	53.75%	1 730	45.29%	336	12.46%	272	7.12%	1 572	58.31%	1 820	47.64%		
	Redis	393	484	0		226	57.51%	251	51.86%	51	12.98%	41	8.47%	238	60.56%	261	53.93%		
	TMux	248	306	0		138	55.65%	141	46.08%	81	32.66%	77	25.16%	169	68.15%	170	55.56%		
C++	Bitcoin	394	521	28	7.11%	264	67.01%	318	61.04%	29	7.36%	27	5.18%	274	69.54%	328	62.96%		
	CodeCompass	27	28	0		12	44.44%	12	42.86%	0		0		12	44.44%	12	42.86%		
	guetzli	72	81	0		44	61.11%	45	55.56%	19	26.39%	17	20.99%	47	65.28%	48	59.26%		
	LLVM-CTE	6 109	6 898	577	9.45%	4 055	66.38%	4 294	62.25%	818	13.39%	801	11.61%	4 160	68.10%	4 399	63.77%		
	OpenCV	3 286	3 714	48	1.46%	2 320	70.60%	2 462	66.29%	533	16.22%	418	11.25%	2 415	73.49%	2 527	68.04%		
	ProtoBuf	313	392	21	6.71%	193	61.66%	214	54.59%	57	18.21%	57	14.54%	201	64.22%	222	56.63%		
	Tesseract	793	880	26	3.28%	461	58.13%	461	52.39%	106	13.37%	96	10.91%	493	62.17%	490	55.68%		
	Xerces	446	532	23	5.16%	303	67.94%	349	65.60%	66	14.80%	64	12.03%	319	71.52%	365	68.61%		
	Z3	2 600	2 801	122	4.69%	2 022	77.77%	2 091	74.65%	861	33.12%	841	30.02%	2 134	82.08%	2 194	78.33%		

# REFERENCES |

- 🌐 andreser. *\_subborrow\_u64 argument order inconsistent.* accessed 2019-12-16. Free Software Foundation - GNU GCC. 2017. URL: [http://gcc.gnu.org/bugzilla/show\\_bug.cgi?id=81294](http://gcc.gnu.org/bugzilla/show_bug.cgi?id=81294).
- 🌐 Mark Capella. *Suspicious code with probably reversed parms in call to IsSingleLineTextControl(bool, uint32\_t).* accessed 2019-11-23. Mozilla. 2016. URL: [http://bugzilla.mozilla.org/show\\_bug.cgi?id=1253534](http://bugzilla.mozilla.org/show_bug.cgi?id=1253534).
- 🌐 Ericsson AB. *CodeChecker – an analyzer tooling, defect database and viewer.* accessed 2020-09-18. 2014. URL: <http://github.com/Ericsson/CodeChecker>.
- 🚩 GrammaTech. *SwapD.* (accessed 2021-04-09). 2019. URL: <http://github.com/GrammaTech/swap-detector>.

## REFERENCES II

-  Hui Liu et al. "Nomen est Omen: Exploring and Exploiting Similarities between Argument and Parameter Names". In: *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. May 2016, pp. 1063–1073. DOI: [10.1145/2884781.2884841](https://doi.org/10.1145/2884781.2884841). URL: <http://ieeexplore.ieee.org/document/7886980>.
-  Michael Pradel and Thomas R. Gross. "Name-Based Analysis of Equally Typed Method Arguments". In: *IEEE Transactions on Software Engineering* 39.8 (Aug. 2013), pp. 1127–1143. ISSN: 2326-3881. DOI: [10.1109/TSE.2013.7](https://doi.org/10.1109/TSE.2013.7). URL: <http://ieeexplore.ieee.org/document/6419711>.

## REFERENCES III

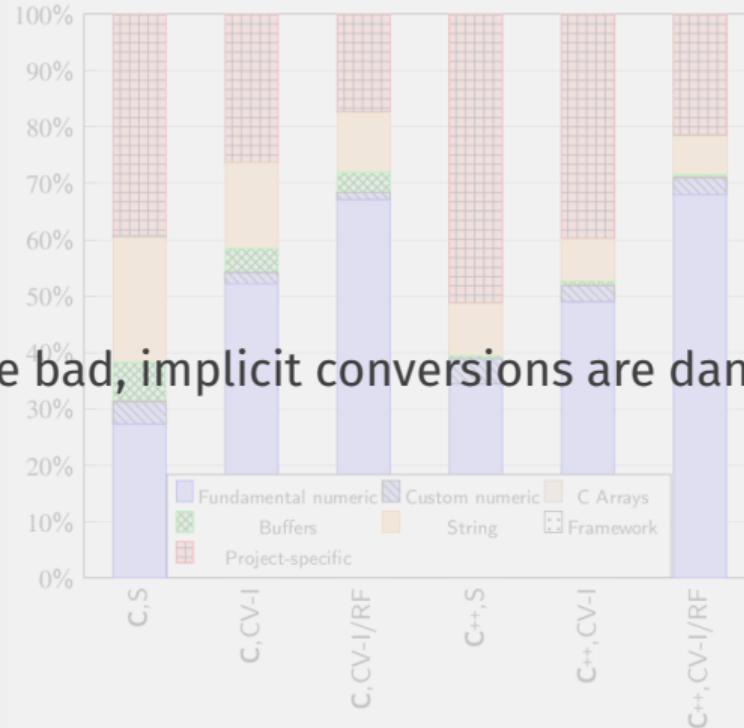
-  Andrew Rice et al. "Detecting Argument Selection Defects". In: *Proceedings of the ACM on Programming Languages* 1.OOPSLA (Oct. 2017), 104:1–104:22. ISSN: 2475-1421. DOI: 10.1145/3133928. URL: <http://doi.acm.org/10.1145/3133928>.
-  Roger Scott et al. "Out of Sight, Out of Place: Detecting and Assessing Swapped Arguments". In: *2020 IEEE 20th International Working Conference on Source Code Analysis and Manipulation (SCAM)*. Sept. 2020, pp. 227–237. DOI: 10.1109/SCAM51674.2020.00031. URL: <http://ieeexplore.ieee.org/document/9252035>.
-  Ray Smith, Google, Inc., et al. *Tesseract OCR Engine*. version 4.1.0 (5280bbc), accessed 2019-12-30. 2006-. URL: <http://github.com/tesseract-ocr/tesseract>.

## REFERENCES IV

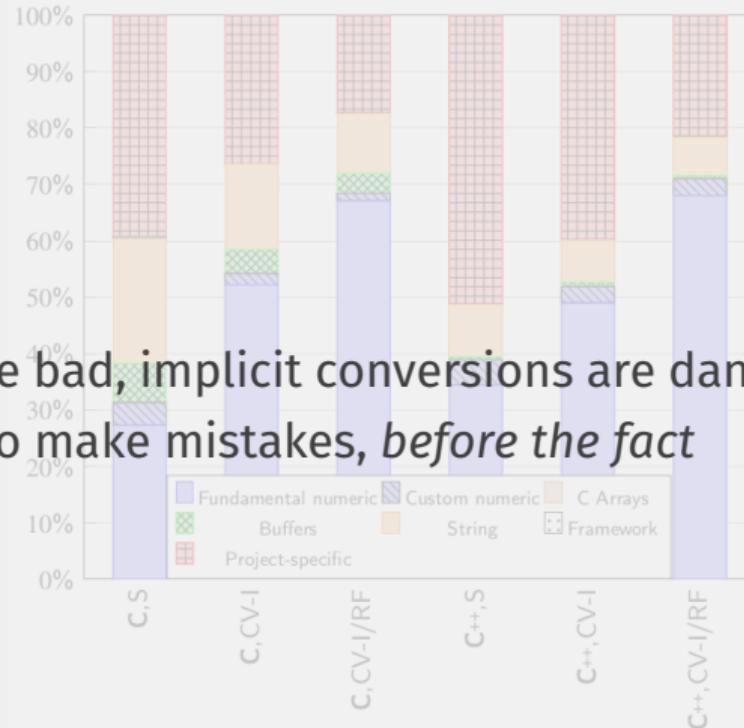
-  The PostgreSQL Global Development Group. *PostgreSQL*. version 12.1 (578a551), accessed 2019-12-30. 1996-. URL: <http://postgresql.org>.
-  Linus Torvalds et al. *git*. version 2.24.1 (53a06cf), accessed 2019-12-30. 2005-. URL: <http://git-scm.org>.
-  János Varjú, Lilla Barancsuk, and Whisperity. *Add readability-suspicious-call-argument check*. accessed 2021-04-13. The LLVM Foundation. 2016. URL: <http://reviews.llvm.org/D20689>.
-  Whisperity. *Add bugprone-easily-swappable-parameters check*. accessed 2021-04-13. The LLVM Foundation. Oct. 2019. URL: <http://reviews.llvm.org/D69560>.

# CONCLUSION

- Built-in types are bad, implicit conversions are dangerous



# CONCLUSION



- Built-in types are bad, implicit conversions are dangerous
- Help users not to make mistakes, *before the fact*

# CONCLUSION

- Built-in types are bad, implicit conversions are dangerous
- Help users not to make mistakes, *before the fact*
- Use tools to find the low-hanging fruits!

