ISO/IEC JTC 1/SC 22/WG23 N0968

Date: 2020-07-06

ISO/IEC TR 24772–10

Notes on this document

This document is an early draft of a Guidance to avoiding programming language vulnerabilities in C++. It started its existence as a direct copy from the equivalent C language document, with the intention to replace the C subclauses with ones that are relevant to C++.

At this point in time, the following clauses are essentially completed first pass.

* 6.2 type system
* 6.3 Bit representation
* 6.4 Floating Point
* 6.5 Enumerator issues [CCB],
* 6.6 Conversion errors
* 6.7 String termination
* 6.8 Buffer boundary violation
* 6.9 Unchecked array indexing
* 6.10 Unchecked array copying (needs to be revisited)
* 6.11 Pointer type conversions
* 6.12 Pointer arithmetic
* 6.13 Null pointer dereference [XYH],
* 6.14 Dangling reference to heap
* 6.15 Arithmetic wrap-around error
* 6.16 Using shift operations for multiplication and division
* 6.17 Choice of clear names [NAI]
* 6.18 Dead Store
* 6.19 Unused variables
* 6.20 Identifier name reuse
* 6.21 Namespace Issues
* 6.22 Initialization of variables [LAV]
* 6.23 Operator precedence and associativity
* 6.24 Side effects and order of evaluation
* 6.25 Likely incorrect expression
* 6.26 Dead store,
* 6.27 Switch statements and static analysis
* 6.28 Demarcation of control flow
* 6.29 Loop control variables
* 6.30 Off-by-one errors
* 6.31 Structured programming
* 6.32 Passing parameters and return values
* 6.33 Dangling references to stack frames
* 6.34 Subprogram signature mismatch
* 6.35 Recursion
* 6.36 Ignored error status and unhandled exceptions
* 6.37 Type breaking reinterpretation of data
* 6.38 Deep vs shallow copying [YAN]
* 6.39 Memory leak and heap fragmentation
* 6.41 Inheritance
* 6.42 Violations of the Liskov substitution principle
* 6.43 Redispatching
* 6.44 Polymorphic variables
* 6.45 Extra intrinsics
* 6.46 Argument passing to library functions
* 6.47 Inter-language calling
* 6.48 Dynamically-linked code and self-modifying code [NYY]
* 6.49 Library Signature
* 6.50 Unanticipated exceptions from library routines
* 6.51 Pre-processor directives
* 6.52 Suppression of language-defined run-time checking
* 6.53 Provision of inherently unsafe operations
* 6.54 Obscure language features
* 6.55 Unspecified behaviour
* 6.56 Undefined behaviour
* 6.57 Implementation-defined behaviour
* 6.58 Deprecated language features
* 6.59 Concurrency -- Activation
* 6.60 Concurrency – Directed termination
* 6.64 Uncontrolled format string

TBD

* 6.2 Type system – issues being fed from 6.40 and elsewhere
* 6.61 Concurrent data access
* 6.62 Concurrency – Premature termination
* 6.63 Protocol lock errors

Participants at meeting 20 July 2020

Stephen Michell

Erhard Ploedereder

Paul Preney

Peter Sommerlad

Richard Corden

Clive Pygott

Michael Wong

Action Items

AI – Richard – In clause 6.2.1 for type system, add text about const. bit-wise vs physical const vs logical const.

AI – Peter, help by Paul – In clause 6.2.1 for an introductory paragraph, write up the introduction to this clause following Erhard’s outline.

AI – Paul – clause 6.2.2, issue about literals that contain specific type information (such as degrees 16C and 16F

AI – Paul - Overriding – 6.43 Private virtual functions can be overridden - write up basic issue. May be a namespace issues or a Beaujolais issue.

AI – Steve – include a comparison of concurrency approaches in clause 4.

AI – Peter – 6.40.2 Templates and Generics - Example needed for the guidance “For binary operator functions, consider providing them as hidden friends”

AI – Peter – Clause 6.64 Format Strings – Provide a better example to show C++ vulnerability

AI – Aaron – In the CERT section on OOP, review the material in 6.2 in this document and provide other applicable rules

AI – Lisa – look at C++ Core Guidelines for “casts” as per 6.2.1 and recommend guidance.

AI – J. Daniel Garcia) – Clause 6.22 Initialization of Variables – Research the Core guidelines for specific guidance on variable initialization for 6.22

AI – Michael Wong – Clause 6.36 Ignored error return and unhahandled exception – Provide a proposal for handling C++ error returns

AI – Michael – Clause 6.59 Concurrency – activation – Verify that the discussion of joinable is complete.

AI - Peter Sommerlad – Clause 6.40 Templates and Generics – An exampole is needed for binary operator functions, consider providing them as hidden friends

 AI – Paul Preney – Clause 6.40 Templates and generics – for the recommendation “ Use qualified-id or this-> to refer to names that may be found in a dependent base” – Need an example and an explanation

AI – Group *– add the guidance from 6.47.2 Interoperability into the Core Guidelines.*

AI – Stephen – Rework Part 1 clause 6.3 to include padding bit issues from 6.47 and 6.64.

Edition 1

ISO/IEC JTC 1/SC 22/WG 23

Secretariat: ANSI

Information Technology — Programming languages — Guidance to avoiding vulnerabilities in programming languages – Part 10 – Vulnerability descriptions for the programming language C++

Document type: International standard

Document subtype: if applicable

Document stage: (10) development stage

Document language: E

*Élément introductif — Élément principal — Partie n: Titre de la partie*

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

**Copyright notice**

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO’s member body in the country of the requester:

*ISO copyright office*

*Case postale 56, CH-1211 Geneva 20*

*Tel. + 41 22 749 01 11*

*Fax + 41 22 749 09 47*

*E-mail copyright@iso.org*

*Web www.iso.org*

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents Page

[Foreword vii](#_Toc1165219)

[Introduction viii](#_Toc1165220)

[1. Scope 1](#_Toc1165221)

[2. Normative references 1](#_Toc1165222)

[3. Terms and definitions, symbols and conventions 1](#_Toc1165223)

[3.1 Terms and definitions 1](#_Toc1165224)

[4. Language concepts 4](#_Toc1165225)

[5. Avoiding programming language vulnerabilities in C++ 4](#_Toc1165226)

[6. Specific Guidance for C++ Vulnerabilities 6](#_Toc1165227)

[6.1 General 6](#_Toc1165228)

[6.2 Type System [IHN] 6](#_Toc1165229)

[6.3 Bit Representations [STR] 7](#_Toc1165230)

[6.4 Floating-point Arithmetic [PLF] 8](#_Toc1165231)

[6.5 Enumerator Issues [CCB] 8](#_Toc1165232)

[6.6 Conversion Errors [FLC] 9](#_Toc1165233)

[6.7 String Termination [CJM] 10](#_Toc1165234)

[ Use std::string or similar, in preference to C-style arrays of chars 11](#_Toc1165235)

[6.8 Buffer Boundary Violation [HCB] 11](#_Toc1165236)

[6.9 Unchecked Array Indexing [XYZ] 12](#_Toc1165237)

[6.10 Unchecked Array Copying [XYW] 13](#_Toc1165238)

[6.11 Pointer Type Conversions [HFC] 13](#_Toc1165239)

[6.12 Pointer Arithmetic [RVG] 15](#_Toc1165240)

[6.13 NULL Pointer Dereference [XYH] 16](#_Toc1165241)

[6.14 Dangling Reference to Heap [XYK] 16](#_Toc1165242)

[6.15 Arithmetic Wrap-around Error [FIF] 18](#_Toc1165243)

[6.16 Using Shift Operations for Multiplication and Division [PIK] 18](#_Toc1165244)

[6.17 Choice of Clear Names [NAI] 19](#_Toc1165245)

[6.18 Dead Store [WXQ] 19](#_Toc1165246)

[6.19 Unused Variable [YZS] 20](#_Toc1165247)

[6.20 Identifier Name Reuse [YOW] 20](#_Toc1165248)

[6.21 Namespace Issues [BJL] 21](#_Toc1165249)

[6.22 Initialization of Variables [LAV] 21](#_Toc1165250)

[6.23 Operator Precedence and Associativity [JCW] 21](#_Toc1165251)

[6.24 Side-effects and Order of Evaluation of Operands [SAM] 21](#_Toc1165252)

[6.25 Likely Incorrect Expression [KOA] 22](#_Toc1165253)

[6.26 Dead and Deactivated Code [XYQ] 24](#_Toc1165254)

[6.27 Switch Statements and Static Analysis [CLL] 24](#_Toc1165255)

[6.28 Demarcation of Control Flow [EOJ] 25](#_Toc1165256)

[6.29 Loop Control Variables [TEX] 26](#_Toc1165257)

[6.30 Off-by-one Error [XZH] 27](#_Toc1165258)

[6.31 Structured Programming [EWD] 28](#_Toc1165259)

[6.32 Passing Parameters and Return Values [CSJ] 28](#_Toc1165260)

[6.33 Dangling References to Stack Frames [DCM] 29](#_Toc1165261)

[6.34 Subprogram Signature Mismatch [OTR] 30](#_Toc1165262)

[6.35 Recursion [GDL] 31](#_Toc1165263)

[6.36 Ignored Error Status and Unhandled Exceptions [OYB] 31](#_Toc1165264)

[6.37 Type-breaking Reinterpretation of Data [AMV] 32](#_Toc1165265)

[6.38 Deep vs. Shallow Copying [YAN] 33](#_Toc1165266)

[6.39 Memory Leak and Heap Fragmentation [XYL] 33](#_Toc1165267)

[6.40 Templates and Generics [SYM] 34](#_Toc1165268)

[6.41 Inheritance [RIP] 35](#_Toc1165269)

[6.41.1 Applicability to language 35](#_Toc1165270)

[6.41.2 Guidance to language users 37](#_Toc1165271)

[6.42 Violations of the Liskov Substitution Principle or the Contract Model [BLP] 37](#_Toc1165272)

[6.42.1 Applicability to language 37](#_Toc1165273)

[6.42.2 Guidance to language users 38](#_Toc1165274)

[6.43 Redispatching [PPH] 38](#_Toc1165275)

[6.43.1 Applicability to language 38](#_Toc1165276)

[6.43.2 Guidance to language users 39](#_Toc1165277)

[6.44 Polymorphic variables [BKK] 39](#_Toc1165278)

[6.44.1 Applicability to language 40](#_Toc1165279)

[6.44.2 Guidance to language users 41](#_Toc1165280)

[6.45 Extra Intrinsics [LRM] 42](#_Toc1165281)

[6.46 Argument Passing to Library Functions [TRJ] 42](#_Toc1165282)

[6.47 Inter-language Calling [DJS] 42](#_Toc1165283)

[6.48 Dynamically-linked Code and Self-modifying Code [NYY] 45](#_Toc1165284)

[6.49 Library Signature [NSQ] 46](#_Toc1165285)

[6.50 Unanticipated Exceptions from Library Routines [HJW] 47](#_Toc1165286)

[6.51 Pre-processor Directives [NMP] 48](#_Toc1165287)

[6.52 Suppression of Language-defined Run-time Checking [MXB] 49](#_Toc1165288)

[6.53 Provision of Inherently Unsafe Operations [SKL] 49](#_Toc1165289)

[6.54 Obscure Language Features [BRS] 49](#_Toc1165290)

[6.55 Unspecified Behaviour [BQF] 50](#_Toc1165291)

[6.56 Undefined Behaviour [EWF] 50](#_Toc1165292)

[6.57 Implementation–defined Behaviour [FAB] 51](#_Toc1165293)

[6.58 Deprecated Language Features [MEM] 52](#_Toc1165294)

[6.59 Concurrency – Activation [CGA] 52](#_Toc1165295)

[6.60 Concurrency – Directed termination [CGT] 52](#_Toc1165296)

[6.60.1 Applicability to language 53](#_Toc1165297)

[6.60.2 Guidance to language users 53](#_Toc1165298)

[6.61 Concurrent Data Access [CGX] 53](#_Toc1165299)

[6.62 Concurrency – Premature Termination [CGS] 53](#_Toc1165300)

[6.63 Protocol Lock Errors [CGM] 53](#_Toc1165301)

[6.64 Uncontrolled Format String [SHL] 54](#_Toc1165302)

[7. Language specific vulnerabilities for C 54](#_Toc1165303)

[8. Implications for standardization 54](#_Toc1165304)

[Bibliography 57](#_Toc1165305)

[Index 60](#_Toc1165306)

# Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard (“state of the art”, for example), it may decide to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 24772-10, was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 22, *Programming languages, their environments and system software interfaces*.

# Introduction

This Technical Report provides guidance for the programming language C++, so that application developers using or considering C++ will be better able to avoid the programming constructs that lead to vulnerabilities in software written in the C++ language and their attendant consequences. This guidance can also be used by developers to select source code evaluation tools that can discover and eliminate some constructs that could lead to vulnerabilities in their software. This report can also be used in comparison with companion Technical Reports and with the language-independent report, TR 24772–1, to select a programming language that provides the appropriate level of confidence that anticipated problems can be avoided.

This technical report part is intended to be used with TR 24772–1, which discusses programming language vulnerabilities in a language independent fashion. It is also intended to be used with TR 24772-3, which discusses how the vulnerabilities introduced in TR 24772-1 are manifested in C, which is a subset of C++.

It should be noted that this Technical Report is inherently incomplete. It is not possible to provide a complete list of programming language vulnerabilities because new weaknesses are discovered continually. Any such report can only describe those that have been found, characterized, and determined to have sufficient probability and consequence.

**Information Technology — Programming Languages — Guidance to avoiding vulnerabilities in programming languages — Vulnerability descriptions for the programming language C++**

# 1. Scope

This Technical Report specifies software programming language vulnerabilities to be avoided in the development of systems where assured behaviour is required for security, safety, mission-critical and business-critical software. In general, this guidance is applicable to the software developed, reviewed, or maintained for any application.

Vulnerabilities described in this Technical Report document the way that the vulnerability described in the language-independent TR 24772–1 are manifested in C++.

# 2. Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14882:2014 — *Programming Languages—C* ++

ISO/IEC TR24772–3 -- Information Technology — Programming Languages — Guidance to avoiding vulnerabilities in programming languages — Vulnerability descriptions for the programming language C

# 3. Terms and definitions, symbols and conventions

## 3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382, in TR 24772–1, in 14882:2014 and the following apply. Other terms are defined where they appear in *italic* type.

The following terms are in alphabetical order, with general topics referencing the relevant specific terms.

3.1.1

abstract

 TBD

3.1.2

 access:

 An execution-time action, to read or modify the value of an object.

Note 1: Where only one of two actions is meant, read or modify. Modify includes the case where the new value being stored is the same as the previous value. Expressions that are not evaluated do not access objects

**3.1.3**

access protection

**alignment**
The requirement that objects of a particular type be located on storage boundaries with addresses that are particular multiples of a byte address.

**3.1.3**

**argument**The expression in the comma-separated list bounded by the parentheses in a function call expression, or a sequence of preprocessing tokens in the comma-separated list bounded by the parentheses in a function-like macro invocation

Note 1: Also called actual argument

Note 2: An argument replaces a *formal parameter* as the call is realized.

**3.1.4**

**behaviour**
An external appearance or action.

Note 3: See: implementation-defined behavior, locale-specific behavior, undefined behavior, unspecified behaviour

**3.1.5**

**bit**
The unit of data storage in the execution environment large enough to hold an object that may have one of two values.

Note 4: It need not be possible to express the address of each individual bit of an object.

**3.1.6**

**byte**
the addressable unit of data storage large enough to hold any member of the basic character set of the execution environment.

Note 5: It is possible to express the address of each individual byte of an object uniquely. A byte is composed of a contiguous sequence of bits, the number of which is implementation-defined. The least significant bit is called the low-order bit; the most significant bit is called the high-order bit.

**3.1.7**

**character**
 An abstract member of a set of elements used for the organization, control, or representation of data.

Note 6: See: single-byte character, multibyte character, wide character

correctly rounded result: The representation in the result format that is nearest in value, subject to the current rounding mode, to what the result would be given unlimited range and precision.

3.1.8

class

TBD

3.1.9

concrete

TBD

3.1.10

diagnostic message

The message belonging to an implementation-defined subset of the implementation’s message output. Note 7: The C Standard requires diagnostic messages for all constraint violations.

3.1.11

dynamic dispatch

TBD

3.1.12

encapsulation

TBD

3.1.13

formal parameter

the object declared as part of a function declaration or definition that acquires a value on entry to the function, or an identifier from the comma-separated list bounded by the parentheses immediately following the macro name in a function-like macro definition.

3.1.14

Implementation

a particular set of software, running in a particular translation environment under particular control options, that performs translation of programs for, and supports execution of functions in, a particular execution environment.

3.1.15

implementation-defined behaviour

the unspecified behaviour where each implementation documents how the choice is made.

Note 8: An example of implementation-defined behaviour is the propagation of the high-order bit when a signed integer is shifted right.

3.1.16

implementation-defined value

an unspecified value where each implementation documents how the choice for the value is selected.

3.1.17

implementation limit

the restriction imposed upon programs by the implementation.

3.1.18

indeterminate value

either an unspecified value or a trap representation.

3.1.19

Inheritance

TBD

3.1.20

language type

see block-structured language, comb-structured language (Non-responsive)

3.1.21

locale-specific behaviour

behaviour that depends on local conventions of nationality, culture, and language that each implementation documents

Note 8: An example, locale-specific behaviour is whether the islower() function returns true for characters other than the 26 lower case Latin letters.

3.1.22

memory location

either an object of scalar[[1]](#footnote-1) type, or a maximal sequence of adjacent bit-fields all having nonzero width

Note 1: A bit-field and an adjacent non-bit-field member are in separate memory locations. The same applies to two bit-fields, if one is declared inside a nested structure declaration and the other is not, or if the two are separated by a zero-length bit-field declaration, or if they are separated by a non-bit-field member declaration. It is not safe to concurrently update two bit-fields in the same structure if all members declared between them are also bit-fields, no matter what the sizes of those intervening bit-fields happen to be. For example a structure declared as

 struct {

 char a;

 int b:5, c:11, :0, d:8;

 struct { int ee:8; } e;

 }

contains four separate memory locations: The member a, and bit-fields d and e.ee are separate memory locations, and can be modified concurrently without interfering with each other. The bit-fields b and c together constitute the fourth memory location. The bit-fields b and c can’t be concurrently modified, but b and a, can be concurrently modified.

3.23

multibyte character

sequence of one or more bytes representing a member of the extended character set of either the source or the execution environment.

Note 9: The extended character set is a superset of the basic character set.

3.1.24

namespace

TBD

3.25

object

region of data storage in the execution environment, the contents of which can represent values.

Note 10: When referenced, an object may be interpreted as having a particular type.

3.1.26

overload

TBD

3.1.27

override

TBD

3.1.28

parameter

(rewrite) See actual argument, argument, formal parameter

3.1.29

Protected

TBD

3.1.30

private

TBD

3.1.31

Public

TBD

3.1.32

Pure

TBD

3.1.33

recommended practice

a specification that is strongly recommended as being in keeping with the intent of the language standard, but that may be impractical for some implementations.

3.1.34

runtime-constraint

a requirement on a program when calling a library function.

3.1.35

single-byte character

the bit representation that fits in a byte.

3.1.36

static

TBD

3.1.37

STL

standard library

3.1.38

template

TBD

3.1.39

trap representation

an object representation that need not represent a value of the object type.

3.1.40

undefined behaviour

the use of a non-portable or erroneous program construct or of erroneous data, for which the language standard imposes no requirements.

Note 11: Undefined behaviour ranges from ignoring the situation completely with unpredictable results, to behaving during translation or program execution in a documented manner characteristic of the environment (with or without the issuance of a diagnostic message), to terminating a translation or execution (with the issuance of a diagnostic message). An example of, undefined behaviour is the behaviour on integer overflow.

3.1.41

unspecified behaviour

the use of an unspecified value, or other behaviour where the language standard provides two or more possibilities and imposes no further requirements on which is chosen in any instance.

Note 12: For example, unspecified behaviour is the order in which the arguments to a function are evaluated.

3.1.42

unspecified value

the valid value of the relevant type where the language standard imposes no requirements on which value is chosen in any instance.

Note 13: An unspecified value cannot be a trap representation.

3.1.43

value

the precise meaning of the contents of an object when interpreted as having a specific type.

Note 14: See implementation-defined value, indeterminate value, unspecified value, trap representation

3.1.44

virtual

TBD

3.1.45

wide character

bit representation capable of representing any character in the current locale.

# 4. Language concepts

*This clause requires a rewrite. See C++ Core Guidelines CPL for a good explanation of the differences*.

C++ has a rich type system with many nuances. In addition to the C base types, int, long, float, double, Boolean, char, and arrays with their C-style vulnerabilities, C++ provides . . .

Many vulnerabilities can be mitigated more easily by using library facilities rather than the base language types. (e.g. std::string rather than char\*)

* Use of the “explicit” keyword for constructors and conversion operators
* operator bool() discussion
* many built-in implicit conversions, refer to TR 24772-3 clause 6.2 and other clauses (C)
* conversion to bool and null pointer conversions
* legacy code operator void\* - change to explicit operator bool
* C-style casts break type safety.
* static\_cast
* explicit casts highlight mismatches between the design and implementation.
* const and volatile
* constexpr – needs a writeup – (in C++:11 , encouraged heavy stack use and possible exhaustion).

Narrowly tailored number-like class types, such as time\_point and duration, improve safety by providing only safe and appropriate operations. User-defined types tailored to a particular use case can provide additional safety.

C++ was initially defined as a syntactic superset of the C programming language: adding object oriented features such as classes, encapsulation, dynamic dispatch, namespaces and templates. It was a “syntactic superset” because whilst there is a core of C++ that is syntactically identical to C, it has always been the case that there are subtle semantic differences between the two, for example:

* Historically, C permitted the use of a function before its declaration (though this is now deprecated in C) . This is illegal in C++
* Where a struct is defined within another struct, in C the inner declaration is in effect made at file scope, so the definition is available for use later in the program. In C++, the inner declaration name is qualified by that of the parent, so without qualification, the inner struct cannot be used later in the program, as in the following example

 struct S1 {

 struct S2 {…} m1;

 …

 };

 struct S2 v1; /\* legal in C not C++ \*/

 S1::S2 v2 // legal in C++ not C

Subsequently, the two languages have diverged, both adding features not present in the other. Not withstanding that, there is still a significant syntactic and semantic overlap between C and C++. So the starting point for this report has been the equivalent for C. However, in many cases, the additional features of C++ provide mechanisms for avoiding the vulnerabilities inherited from C, and these are reflected in the following sections.

*Include discussions of Object orientation,* ***static****, and* ***const,*** *scoped enumerations*

# 5. Avoiding programming language vulnerabilities in C++

In addition to the generic programming rules from ISO/IECTR 24772-1 clause 5.4, additional rules from this section apply specifically to the C++ programming language. The recommendations of this section are restatements of recommendations from clause 6, but represent ones stated frequently, or that are considered as particularly noteworthy by the authors. Clause 6 of this document contains the full set of recommendations, as well as explanations of the problems that led to the recommendations made.

Every guidance provided in this section, and in the corresponding Part section, is supported by material in Clause 6 of this document, as well as other important recommendations.

***TBD***

|  |  |  |
| --- | --- | --- |
| Index |  | Reference |
| 1 |  |  |
| 2 |   |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |

Need to consider C++-11, 14 and 17.

# 6. Specific Guidance for C++ Vulnerabilities

## 6.1 General

This clause contains specific advice for C++ about the possible presence of vulnerabilities as described in TR 24772-1, and provides specific guidance on how to avoid them in C++ code. This section mirrors TR 24772-1 clause 6 in that the vulnerability “Type System [IHN]” is found in 6.2 of TR 24772–1, and C++ specific guidance is found in clause 6.2 and subclauses in this TR.

As part of its design (and with few exceptions), C++ has a common subset with the complete C language. For code portions written in the common subset, the vulnerabilities described and the advice given in ISO/IEC TR 24772-3:2020, *Part 3 – Vulnerability descriptions for the programming language C*,  apply, except when this document provides refined advice. The following subclauses usually do not further acknowledge the issues from the subset since those have been adequately addressed in the referenced document. However, C++ provides mechanisms to mitigate many of the problems that arise. Please refer to the respective clauses of this document for these mitigations and related guidelines.

## 6.2 Type System [IHN]

### 6.2.1 Applicability to language

C++ is a strongly- and statically-typed language: all variables and expressions must have a type. C++ also permits implicit and explicit conversions between types.

*Erhard proposes the following:*

*What should be here are:*

*Static or dynamic type checks? (watch out: in templates?)*

 *type equivalence model (which types are implicitly convertible/promotable).*

 *risks of structural equivalence*

 *A few high-level words about explicit casts/promotions, plus link to 6.6 and 6.37.*

 *How do I keep two semantic types with identical representation apart? Celsius and Fahrenheit (via named typedefs?)*

 *Can I encode ranges (sort of, implicitly by int8, int16, etc.)*

AI – Peter, help by Paul – Write up the introduction to this clause following Erhard’s outline.

C++-style casts eliminate many of the issues associated with C-style casts. See clauses 6.6 and 6.37 for a discussion of C++ conversion capabilities and errors.

Holes in the type system

Shadow type system (explain) – Thoughts. – exception specification don’t participate in a functions types, until they are invoked. (Michael to provide some text in an email). This is only an issue before C++ 17. Fixed by using noexecpt and avoiding throw.

Don't write exception specifications on your functions unless you're forced to (because other code you can't change has already introduced them and/or you are in pre C++17 when dynamic exception was banned or C++20 when empty throw specification was banned to be replaced by noexcept)

A specific issue exists with exceptions in that exception specifications aren't part of a function's type, except when they are. They form a shadow type system whereby writing an exception specification can be:

* Illegal: In a typedef for a pointer to function.
* Allowed: In the identical code without the typedef.
* Required: In the declaration of a virtual function that overrides a base class virtual function that has an exception specification.
* Implicit and automatic: In the declaration of the constructors, assignment operators, and destructors when they are implicitly generated by the compiler.

A common but incorrect belief is that exception specifications statically guarantee that functions will throw only listed exceptions (possibly none), and will enable compiler optimizations based on that knowledge. Instead, exceptions cause the compiler to inject additional run-time overhead in the form of implicit try/catch blocks around the function body to enforce via run-time checking that the function does in fact emit only listed exceptions (possibly none), unless the compiler can statically prove that the exception specification can never be violated in which case it is free to optimize the checking away. In addition, exception specifications can both enable and prevent further compiler optimizations (besides the inherent overhead already described); for example, some compilers refuse to inline functions that have exception specifications.

When these rules are violated, by default they immediately terminate your program.

You generally can't write useful exception specifications for function templates anyway, because you generally can't tell what exceptions the types they operate on might throw.

 Exceptions : - see 6.36

 Incomplete types and operations on incomplete types – usually checked by the compiler but some holes exist, for example when a virtual function is invoked from within a class constructor or destructor.

 Mitigation, don’t overload unary ‘&’ or “,”, don’t delete an object of an incomplete class type,

AI – Paul – explain.

 don’t invoke virtual functions in constructors and destructors

High level issues

Note: discuss the problems with unions.

*NOTE: Ensure that the above explicit text is put into the relevant sections and generalize or forward reference in this section.*

C++ overloading of operators can cause significant issues. See clause 6.23.

One hazard is that the overloaded operators do not automatically connect the inverse operator, such as {==, !=}, {<, >=}, and {>, <=}. Unless the declarer declares all relevant operators, unexpected results can occur. In addition, overloaded operators && and || do not have shortcut semantics and thus behave differently than the default operators. This is problematic since the safety of a right-hand operand of a short circuit operator often depends on the result of the left-hand operand, e.g. an existence test before the value is read.

AI – 66-1 Richard – add text about const. bit-wise vs physical const vs logical const.

.

The primitive numeric types of C++, for historical reasons, allow a variety of implicit conversions, some of which are unsafe. C++ class types, in contrast, have strictly limited implicit operations and conversions, and may practically be used in place of primitive numeric types.

*13 Feb 2020 - Issue moved here from 6.41(?) Templates and Generics- Templates take a hit because it is possible to invoke a template with a type that does not have all of the operations necessary to correctly execute the operations. For example, if an instantiation of a template type parameter does not implement “<” (or is it “less”?) and the underlying code executes a “sort”, the sort will call whatever “<” it can use. If a “<” had been defined for a related type, then that “<” will be called. The experts at meeting 67 say that this is not exclusive to templates, so the problem should be assigned here and referenced from templates.*

*13 Feb 2020 – Another issue appears to be that some primitive types have operations that are inappropriate for the type. For example, the basic type for characters include operators “+”, “-“, “<”, “\*” which operate on the underlying integer representation. Passing such types into templates or into a class can result in the fundamental integer operation being called instead of a higher-level operation to perform, for example, comparison on a ISO 8859 character set. The only work-around appears to be to define classes with the fundamental type and legal operations as the only operators.*

*Question – how does the “.” Operator interact with such types?(Maybe this comma operator?)*

*A recommendation for future revisions of the language could be to provide a way to remove basic operations from types such as enumeration types or characters.*

References

* CERT section OOP (AI – Aaron to provide others), (note that some of these will likely migrate to other vulnerabilities)x
	+ DCL52-CPP. Never qualify a reference type with const or volatile
	(this one is odd because the language makes this an error, but some
	compilers like MSVC only warn on it, but will still translate the
	source somehow)
	+ EXP55-CPP. Do not access a cv-qualified object through a cv-unqualified type
	+ EXP47-C. Do not call va\_arg with an argument of the incorrect type
	+ OOP51-CPP. Do not slice derived objects
* AI – Lisa – look at C++ Core Guidelines for “casts” (-> 6.6)
* AUTOSAR (AI Peter to work with AUTOSAR to provide references)

### 6.2.2 Guidance to language users

For specific types discussed in this document, such as floating point types, see the respective clauses.

* Be aware of the rules for typing and conversions with fundamental types (i.e., built-in language types) and operators to avoid vulnerabilities.
	+ To help aid correctness of code, couple all scalar values (e.g., double, complex, int, etc.) that have units (e.g., metres, grams, litres, etc.) with suitable types representing those units. This will allow the compiler to generate errors with scalar-with-unit-type-values when they are used with operations that are incompatible.
	+ User-defined literals (This lets you define 15C and 15F)
		- (AI – Paul – Write up)
* Make converting constructors and conversion functions explicit to avoid them being used implicitly or in unexpected ways. (*Move to 6.6)*
* Do not use union without an associated indicator as to which member it holds, alternatively wrap a union in a class with a type field or use C++:17 std::variant (see clause 23.7 of ISO/IEC 14882:2017)
* Use unsigned integer types when manipulating integral values' bits.
* Use one of (not both!) unsigned or signed integral types within expressions to help avoid programming mistakes, consistently
* Treat explicit casts as candidates for code refactoring, i.e., ideally explicit casts should not be required in the code.
* Avoid casting away const:. Doing so can result in undefined behaviour that may not be detectable by the compiler or other tools. Refactor code so that it handles const and non-const types properly.
* don’t overload unary ‘&’ or “,”,
* don’t delete an incomplete class type,
* don’t invoke virtual functions in constructors and destructors

[+ the guidance already in the document as WG23 decides (or moves to another section or otherwise edits) --there are a lot of items listed]

From Part 1. The Part 1 guidelines were accepted but are left here for review.

* Take advantage of any facility offered by the programming language to declare distinct types and use any mechanism provided by the language processor and related tools to check for or enforce type compatibility.
* Use available language and tools facilities to preclude or detect the occurrence of implicit type conversions, such as those in mixed type arithmetic. If it is not possible, use human review to assist in searching for implicit conversions.
* Avoid explicit type conversion of data values except when there is no alternative. Document such occurrences so that the justification is made available to maintainers.
* Use the most restricted data type that suffices to accomplish the job. For example, use an enumeration type to select from a limited set of choices (such as, a switch statement or the discriminant of a union type) rather than a more general type, such as integer. This will make it possible for tooling to check if all possible choices have been covered.
* Always respect the implied unit systems, when converting explicitly from one numeric type to another.

(Explicit C++ guidance for unit-based types.

* Follow the guidance of TR 24772-1 clause 6.2.5.
* Use distinct C++ types for unit systems if available or define explicit unit-based types.)
* Treat every compiler, tool, or run-time diagnostic concerning type compatibility as a serious issue. Do not resolve the problem by modifying the code to include an explicit conversion, without further analysis; instead examine the underlying design to determine if the type error is a symptom of a deeper problem.
* *Never ignore instances of implicit type conversion; if the conversion is necessary, change it to an explicit conversion and document the rationale for use by maintainers. – narrowing conversions and loss of precision*
* Analyze the problem to be solved to learn the magnitudes and/or the precisions of the quantities needed as auxiliary variables, partial results and final results.
* Create types that more accurately model the problem domain, with corresponding safe operations and conversions in lieu of using primitive types.
* Minimize use of predefined numeric types whose ranges and precisions are implementation defined. Instead, use types whose ranges and precision are guaranteed.
* *C++ Issue – Use syntax that forces the compiler to Issue diagnostics on narrowing – need example.*
* Treat every explicit cast as a candidate for refactoring.
* Use C++ casts rather than C-style casts, as they provide more compile-time checking and are more restrictive in what they can change, - rationale – syntactic distinction – in C++ obvious.
* *Make class member functions that can be static, ‘static’. Make class member functions that cannot be ‘static’, but can be ‘const’, ‘const’*
* *The ‘mutable’ keyword for class member variables should be used sparingly*
* Don't mix signed and unsigned types in arithmetic
* Follow the advice provided in TR 24772-3 clause 6.2.2. when using C-style numeric types, and implicit conversions.

## 6.3 Bit Representations [STR]

### 6.3.1 Applicability to language

This vulnerabilities described in TR24772-1 clause 6.3 is applicable to C++.

*Document the C++ behaviours- handling bit-fields, - hitting enclosing word, concurrent access, hardware implications,*

Able to use non-integer types (such as enumerations) in accessing bit fields.

A C++ memory location is either an object is or a contiguous collection of bit-fields.

C++ bit fields are not separated from adjacent bit-fields for purposes of thread synchronization or volatility. Bit-fields are very difficult to use correctly in these contexts.

TO-DO – Bit\_Cast - Applies a bit representation to an object of a new type, for example copy a 64 bit IEEE 654 number to a 64 bit integer to perform bit operations and then cast back.

Needs documentation.

When a struct, union or class is embedded within an array, implementations will typically add padding to provide efficient alignment and access. Therefore the compiler will add padding bytes in case it is used in arrays. The padding bytes can be used as a *secret channel*

**6.3.2 Guidance to language users**

In addition to the advice of TR 24772-3 clause 6.3.2:

See C++ Core Guidelines ES101 use unsigned types for bit manipulation.

CERT INT34-C

* Do not use std::vector<bool>
* Use bit-fields with care or avoid them entirely. Instead, use a class type containing one or more unsigned integer data members and member functions appropriate to the particular situation.
* Do not create a bit-field of a signed type and size one.

See AUTOSAR A9-6-1

* Avoid using bit\_cast or heavily analyse and document its use.
* Prefer performing member-by-member copies and moves instead of using std::memcpy or std::memmove
* Cover known padding with named members to eliminate padding. This can be checked with a static\_assert( std::has\_unique\_object\_representations\_v<T>). If this assertion is true then there are no padding bits.

CERT EXP62-CPP

## 6.4 Floating-point Arithmetic [PLF]

### 6.4.1 Applicability to language

C++ has the vulnerability as described in ISO/IEC TR 24772-1 clause 6.4. The C++ standard assumes IEC 60559 if std::numeric\_limits<T>::is\_iec559 is true for the types in use. In the absence of this, C++ makes few guarantees about the behaviour of floating point numbers.

Standard library comparison functions default to the predefined comparisons of floating point types which can produce surprising results due to the properties of floating point. See clause 6.40 Templates and Generics.

*<pick up generl statement about NaNs and zero and -zero>Issues with std::less and std::equal affect programmers because a common paradigm is to sort containers of floating point numbers. Sort invokes std::less which can give erroneous results.*

### 6.4.2 Guidance to language users

* Follow the general advice of ISO/IEC TR 24772-1:2019 clause 6.4.5
* Verify compliance to ISO/IEC/IEEE 60559:2011 at compile time through std::numeric\_limits<T>::is\_iec559. Other numeric characteristics such as min(), max(), existence of NaNs, has\_denorm, and infinities can be determined in this class template.
* Be aware that the default comparison functions in the standard library may produce wrong results when used on floating point members. In particular std::less is not a total order; std::equal is not equivalent to substitutability (NaNs compare unequal to themselves, but neither less nor greater, and negative zero compares equal to positive zero)

## 6.5 Enumerator Issues [CCB]

### 6.5.1 Applicability to language

### 6.5.1.1 References

AUTOSAR A7-2-2 Enumeration base type shall be explicitly defined

6.5.1.2 **Applicability**

C++ offers enums for defining distinct types composed of sets of related named constants. The type of each enum is different from all other types. Each enum has an underlying integral type, which the user can specify. Since enums are distinct types, the user can only assign values to an object of enumerated type that are values of that enumerated type. C++ does not support implicit conversion of an int to an enum, therefore preventing A = B + C where A, B and C are variables of the same enum, unless an overloaded operator “+” is provided.

C++ enums can be scoped (enum class) or unscoped (enum). C++ supports implicit conversion of an unscoped enum to an integer by integral promotion

enum Color : short {red, green, blue};

short i = red; // implicit conversion

C++ does not support implicit conversion of a scoped enum to an int. Hence, operations such as ++, +, < and enums used as array indices require explicit definitions.

enum class Color : short {red, green, blue};

short i = red; // error – no implicit conversion

Where unscoped enums are used as array indexes and have a user-specified mapping to an underlying representation, there will be “holes” as documented in TR24772-1 clause 6.6.

Note that unscoped enumeration types implicitly promote their underlying type and can be used as the index of an array without a cast, with all of the issues described in TR 24772-1 clause 6.5.

From C++ 2017 forward, casting a value to an enumeration type is undefined behavior unless the source value is within the range of values of an enumeration type. See CERT INT50-CPP.

### 6.5.2 Guidance to language users

* Use *scoped enumerations* in preference tothe C-style *unscoped enumerations* for related values, especially at namespace-level.
	+ See CPP Core Guidelines Enum.3 “Prefer class enums over ‘plain’ enums”.
	+ See AUTOSAR A7-2-3 “Enumerations shall be declared as scoped enum classes”
	+ See MISRA C++ 28.5.5
* Use constexpr to declare a set of unrelated values, such as
constexpr size\_t bufferLen = 128;
constexpr char special\_char = ‘a’;
* Provide operators and functions that perform the arithmetic operations and conversions appropriate to the enumerated type. Outside those functions, avoid directly performing arithmetic or conversions on objects of the enumerated type.
	+ See CPP Core Guidelines Enum.4 “Define operations on enumerations for safe and simple use”
* If *unscoped enumerations* are used, follow the general advice of TR 24772-3 clause 6.5.2 as well as the following:
* Avoid casting arbitrary integer values to enumeration type. If it is unavoidable, use braced initialization instead of C-style or static casts
 e\_type{7};

See CERT INT50-CPP “Do no Cast to an out-of-range-value”

* Obtain the underlying enumeration value, by casting the enumeration to its underlying type, e.g.,

enum e\_type{A, B, C};

auto value = static\_cast<std::underlying\_type\_t<e\_type>>(B);

## 6.6 Conversion Errors [FLC]

### 6.6.1 Applicability to language

C++ includes some of the conversion mechanisms of C, as documented in TR 24772-3 clause 6.6.1.

C++ type conversion mechanisms differ from the mechanisms of C, as documented in ISO IEC 14882 Annex C. This subclause highlights those differences where C++ eliminates potential vulnerabilities found in C.

Implicit, i.e., automatic, conversions to a type T can be performed, for example, in the following situations:

1. If the declaration, T t=e;, is defined for some expression, e, and some invented variable, t [C++17, Clause 7 [conv], para 3];
2. In expressions involving operands of operators (e.g., +, -, \*, /, etc.) subject to the requirements of each operators' operands [C++17, Clause 7 [conv], para 2.1];
3. For example, the expression, 5 + 6.5, has operands of type int and double. Per language rules, the int will be implicitly converted to double, i.e., the expression becomes double(5) + 6.5, i.e., 5.0 + 6.5.
4. In the condition of an if, for, do..while, or while statement: the implicit conversion will be to the type bool [C++17, Clause 7 [conv], para 2.2];
5. In the expression of a switch statement: the implicit conversion will be to an integral type [C++17, Clause 7 [conv], para 2.3];
6. In an expression that initializes an object (e.g., an argument to a function call, the expression in a return statement) [C++17, Clause 7 [conv], para 2.4];
7. When a non-explicit class/struct/union constructor can be invoked on an object resulting in some desired type, T, from initial objects passed to the constructor; and
8. When a conversion operator has not been declared explicit*,* it can be implicitly invoked on an object resulting in some desired type, T, from an initial type.

Explicit conversions are conversions that occur:

 From the C++ reference manual clause 8.5.3 paragraph 2:

“An explicit type conversion can be expressed using functional notation (8.5.1.3), a type conversion operator (dynamic\_cast, static\_cast, reinterpret\_cast, const\_cast), or the cast notation.”

Note: The final 'cast notation' refers to C-style cast.

Implicit conversions from void\* to any other object type is invalid.

C++ adds a number of new features relevant to type conversion:

* C-style casts (using the desired type in brackets in front of an expression), whilst still available in C++, are augmented by four C++ specific cast and function style casts. These provide a number of (mostly) compile-time checks, so prevent casting between obviously inappropriate types
* The programmer can add code to the definition of a class to allow values of any other type to be implicitly cast to that class type, or for a class object to be implicitly cast to any other type (including basic numeric types). As implicit conversions can make code maintenance more difficult, in general they should be avoided

Non-explicit unary constructors, non-explicit conversion operators, *(and conditionally-explicit unary constructors(??))* can provide implicit conversions that are unexpected by the programmer. For such constructors and conversion operators should be declared with the keyword **explicit**.

Implicit casting to a class type occurs when a class has a constructor that can take a single parameter, as in the following example:

class C

 {public:

 C(int x=10, float y=0){…}

 };

void foo(C param){…}

… foo(21); …

The call to foo requires a parameter of type C, but is provided with an int. However, as C has a constructor that can take an int parameter (the float parameter is ignored because it has a default value), a temporary object of type C is constructed using 21 as the x parameter. This is passed to foo. The temporary object is destroyed when foo returns.

Note that this implicit conversion to a class object is the default behavior of constructors that can be called with a single parameter. To prevent this happening, the keyword ‘explicit’ is used before the constructor, as in:

 explicit C(int x=10, float y=0){…}

The call foo(21) would now not be legal.

C++ provides:

* static\_cast (explain)
* dynamic\_cast (explain)
* const\_cast(explain)
* reinterpret\_cast (as in <target\_type>(expression)) that casts an arbitrary piece of data to the desired type.

Unlike C++'s other cast notations, dynamic\_cast relies on run-time type information generated by the compiler to ensure the requested conversion is valid. If it is not valid, then the null pointer value of the required result type is returned for pointer types, otherwise an exception is thrown. [C++17, Clause 8.2.7 [expr.dynamic.cast]] Thus, dynamic\_cast is safer to use when converting down a hierarchy.

All other conversions are not necessarily "safe" as they can sometimes yield unexpected results. This is an issue with implicit conversions since they are automatic: the programmer does not explicitly write code to do the conversion. For example, a common problem is mixing signed and unsigned integral types in arithmetic expressions. This can become a problem since the ranges of signed and unsigned integer types differ; unsigned integers are permitted to overflow and wrap (modulo arithmetic) whereas signed integers should never overflow. When these types are mixed in the same expressions, there is a possibility that erroneous values will result. (example? AI - Paul)

### 6.6.2 Guidance to language users

In addition to the general advice of TR 24772-1 clause 6.6.5:

* Guidance for numeric conversions: Use the brace form of function style casts
* Use C++ casts rather than C-style casts, as they provide more checking
* If a class has a converting constructor and implicit conversions are not required, make that constructor ‘explicit’
* Avoid casts (see Core guidelines ES48)
* If using a cast, use a named cast (see C++ Core Guidelines ES49)
* Don’t cast away const (see C++ Core Guidelines ES50)
* Prefer {} over () when constructing values. (See C++ Core Guidelines ES23 and ES64)
* Avoid narrowing conversions. (See C++ Core Guidelines ES46)
* Don’t mix signed and unsigned arithmetic (See C++ Core Guidelines ES100)

*Ensure that the rationale for this guidance is addressed above*

## 6.7 String Termination [CJM]

### 6.7.1 Applicability to language

The vulnerability as documented in TR 24772-1 exists in C++ when C-style strings are used. A string in C++ is composed of a contiguous sequence of characters terminated by and including a null character (a byte with all bits set to 0). Therefore strings in C++ cannot contain the null character except as the terminating character. Inserting a null character in a string either through a bug or through malicious action can truncate a string unexpectedly. Alternatively, not putting a null character terminator in a string can cause actions such as string copies to continue well beyond the end of the expected string. Overflowing a string buffer through the intentional lack of a null terminating character can be used to expose information or to execute malicious code.

C++ provide a string class (in the iostream library), std::string. Internally, the class maintains an array of char on the heap. If an attempt is made to copy or append a string that results in a string larger than the current size of the array, a new larger array is allocated.

UNICODE and multibyte strings??

### 6.7.2 Guidance to language users

* Use std::string or similar, in preference to C-style arrays of chars
* Provide guidance on collecting C-style strings at nterfaces and converting them to std::string.

## 6.8 Buffer Boundary Violation [HCB]

### 6.8.1 Applicability to language

A buffer boundary violation condition occurs when an array is indexed outside its bounds, or pointer arithmetic results in an access to storage that occurs outside the bounds of the object accessed. This behaviour may occur when copying, initializing, writing or reading.

In C++, the built-in subscript operator [] is defined such that E1[E2] is identical to (\*((E1)+(E2))), so that in either representation, the value in location (E1+E2) is returned. C++ does not perform bounds checking on arrays: arrays may be accessed outside of their bounds which is undefined behaviour. For example, in C++ the following code is syntactically valid, though, if offset has the value 10, the behaviour is undefined:

int foo(const int offset) {

 int t;

 int x[] = {0,0,0,0,0};

 t = x[offset];

 return t;

 }

or, when written using iterators, the same issues can occur

int foo(const int offset) {

  std::array<int, 5> a;

  return \*(a.begin() + offset);

For further explanation and examples, see

[https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR50-CPP.+Guarantee+that+container+indices+and+iterators+are+within+the+valid+range](https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR50-CPP.%2BGuarantee%2Bthat%2Bcontainer%2Bindices%2Band%2Biterators%2Bare%2Bwithin%2Bthe%2Bvalid%2Brange)

[https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR53-CPP.+Use+valid+iterator+ranges](https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR53-CPP.%2BUse%2Bvalid%2Biterator%2Branges)

[https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR55-CPP.+Do+not+use+an+additive+operator+on+an+iterator+if+the+result+would+overflow](https://wiki.sei.cmu.edu/confluence/display/cplusplus/CTR55-CPP.%2BDo%2Bnot%2Buse%2Ban%2Badditive%2Boperator%2Bon%2Ban%2Biterator%2Bif%2Bthe%2Bresult%2Bwould%2Boverflow)

Note: Consider C++ Core guidelines if completed.

*As described in 6.7 [CJM], C++ provides library functions, e.g. std::string, that encapsulate strings and prevent boundary violations when accessing arrays of characters. It also provides standard templates that provide similar facilities for any other type, such as std::vector.*

### 6.8.2 Guidance to language users

* Avoid C-style arrays. Guidance for the use of C-style arrays is provided in TR 24772-3 clause 6.8.2.
* Use a library class such as std::array to encapsulate an array, or write a class with similar behavior.
* Use library classes such as gsl::span or std::string\_view to represent ranges of elements within an array or container.
* Use containers of the standard library, such as std::vector or std::deque, to model arrays with dynamically changing size.
* Use iterator-based algorithms, such as those of the standard library.
* Use the range-based for loop construct such as for (auto I: *some container*) to iterate within the defined bounds of the object.
* Use iterators over the range of elements to be accessed instead of using an array and bounds as parameters.
* Perform range checking before indexing into an array. In the interest of speed and efficiency, range checking only needs to be done when it cannot be statically shown that an access outside of the array cannot occur.
* When performing random access by indexing, follow the guidance of clause 6.9.2. When performing other forms of random access, follow the guidance of clause 6.12.2. *)*
* Use static analysis tools to detect buffer boundary violations.

## 6.9 Unchecked Array Indexing [XYZ]

### 6.9.1 Applicability to language

Placing C-types (such as arrays) in containers let the implementer create access operators that check bounds.

Like a C-style array, some STL containers, such as std::vector, can be indexed using [], and as in C such an access is unchecked. However, these containers also provide an access function at() that behaves like [], but performs a check that the access is within the bounds of the container.

Similar issues arise from accessing elements in containers by pointer arithmetic.

The following example compares C and C++ performing equivalent array operations:

|  |  |  |
| --- | --- | --- |
| **C** | **C++** | **Comment** |
|  | #include <array> |  |
| int arr [10]; | std::array<int,10>arr; | Both arrays are of 10 elements |
| arr[10] = 0; | arr[10] = 0; | Both accesses silently violate array’s bounds |
| arr[10] = 0; | arr.at(10) = 0; | The C++ access fails with an error exception |

6.9.2 Guidance to language users

* Follow the guidance from clause 6.8.2.
* Use static analysis or explicit checks to establish that bounds violations do not occur. Otherwise use the at() member function of the standard library containers and handle the bounds violation exceptions. See clause 6.36 Ignored error status and unhandled exceptions.

## 6.10 Unchecked Array Copying [XYW]

### 6.10.1 Applicability to language

This subclause requires a complete rewrite.

A buffer overflow occurs when some number of bytes (or other units of storage) is copied from one buffer to another and the amount being copied is greater than is allocated for the destination buffer. In essence this is a special case of Buffer Boundary Violation [HCB].

As with [HCB], in most cases the vulnerability can be avoided by using library classes, such as std::vector, which provides a copy assignment operator that adjusts the size of the target to fit the object being copied.

If for some reason this is not acceptable, C++ has access to the C library functions memcpy and memmove. Both simply copy memory and no checks are made as to whether the destination area is large enough to accommodate the amount of data being copied. It is assumed that the calling routine or programmer has ensured that adequate space has been provided in the destination. Problems can arise when the destination buffer is too small to receive the amount of data being copied.

In general, placing C-types (such as arrays) in containers lets the implementer create whole array operations that can eliminate the errors discussed in Part 1, clause 6.10.

### 6.10.2 Guidance to language users

This subclause requires a complete rewrite.

* Use standard library containers, such as std::vector, that provide copying mechanisms that ensure the target array is large enough for the indicated source.
* For copies of fixed-sized arrays, perform range checking to prevent out-of-bounds access on the target and the source arrays. In the interest of speed and efficiency, range checking only needs to be done when it cannot be statically shown that an access outside of the arrays cannot occur.
* Use std::string\_view to represent immutable string literals.
* Use std:string to represent mutable strings.
* Place arrays in containers with whole-array and bounds-checking operations.

## 6.11 Pointer Type Conversions [HFC]

### 6.11.1 Applicability to language

In this clause, all C++ references, in addition to pointers. The shared\_ptr casts

The vulnerabilites as described in TR 24772-1 clause 6.11.1 also apply to C++.

In general casting pointers breaks the type system and should be avoided.

In C++, a C-style cast is defined in terms of the C++ cast operators const\_cast, static\_cast, and reinterpret\_cast. In some cases, it is unspecified which cast is used, for example when a cast operation involves an incomplete type, a reinterpret\_cast may be used for the conversion which can produce an incorrect result.

Reinterpret\_cast has the problem that it simply treats the unmodified pattern of bits in the pointer as being of the target type rather than the original type, but the C++ standard recognizes that the language or compiler may impose constraints or additional data requirements on a pointer. Static\_cast and dynamic\_cast take this difference into account, but other cast operators do not take this into consideration and hence can give incorrect results. For example, in the use of multiple inheritance, the address of an object may be different than one of its base class sub-objects, causing the potential for the exploitable access of adjacent memory.

C++ permits the change of constant or volatile properties as part of a conversion. Such conversions, unless done in extremely limited ways, puts the program at risk of creating undefined behavior.

A typical use of pointer conversion in C++ is where there is a hierarchy of classes declared, as in:

 struct Base {virtual ~Base() = default; };

 struct Derived: Base { };

Where a Base pointer needs to be converted to Derived pointer, dynamic\_cast will check at runtime that the pointer is to an object of the correct type. If it’s not, either nullptr will be returned, or an error exception thrown.

C++ Dynamic cast and the use of it during construction and deconstruction needs further exposition. The this pointer type can have surprising effects.

AI – Paul, Richard to review – Writeup about incomplete objects before or after their lifetime and related to translation units.

Pointer casts to a more strictly aligned pointer type is undefined behaviour.

Reinterpret\_cast for pointer-interconvertible on objects (see clause 6.9.2 of IS 14882)

C++ permits reinterpret\_cast to be used to convert a pointer to an object, a, to a pointer to another object, b, only in specific restricted circumstances, i.e., when

* a and b are the same object,
* either a or b is a standard-layout union object and the other is a non-static data member of that object,
	+ Examples:
	union A { int i; double d; } a;
	int\* iptr = reinterpret\_cast<int\*>(&a);
	double\* dptr = reinterpret\_cast<double\*>(&a);
	A\* uptr1 = reinterpret\_cast<A\*>(iptr);
	A\* uptr2 = reinterpret\_cast<A\*>(dptr);
* either a or b is a standard-layout class object and the other is the first non-static data member of that object,
	+ Examples:
	struct B { int i; double d; } b;
	int\* iptr = reinterpret\_cast<int\*>(&b);
	B\* bptr = reinterpret\_cast<B\*>(iptr);
* either a or b is a standard-layout class object with no non-static data members and the other is the first base class subobject of that object, or,
	+ Examples:
	struct A { double d; };
	struct B : A { static int i; } b;
	double\* dptr = reinterpret\_cast<double\*>(&b.d);
	B\* cptr = reinterpret\_cast<B\*>(dptr);
* there exists an object c where a and c are pointer-interconvertible and c and b are pointer-interconvertible.

In essence, such pointer-interconvertibility implies objects a and b have the same address, however, having the same address does not imply a and b are pointer-interconvertible! For example, an array and its first element have the same address but they are not pointer-interconvertible. This means that one cannot use reinterpret\_cast to cast an array object to the type of its first element or vice versa. [Reference: ISO 14882 Section 6.9.2 [basic.compound], Paragraph 4].

### 6.11.2 Guidance to language users

* Follow the advice provided by TR 24772-1 clause 6.11.5.
* Avoid the C-style cast, reinterpret\_cast, and casts to and from void\*.
* For conversions that remove the constant qualification, see the guidance in TR24772-1 clause 8.2.5
* or volatile qualifications
* When downcasting, prefer dynamic\_cast and explicitly handle the possible failure cases.
* References???
* Heed compiler warnings that are issued for pointer conversion instances. The decision may be made to avoid all conversions so any warnings must be addressed. Note that casting into and out of void \* pointers will most likely not generate a compiler warning as this is valid in C++
* Use new and delete to allocate/deallocate memory, rather than malloc/free.
	+ EXP51-CPP. Do not delete an array through a pointer of the incorrect type (implied by 6.11)
	+ EXP57-CPP. Do not cast or delete pointers to incomplete classes (-> 6.11)
	+ EXP36-C. Do not cast pointers into more strictly aligned pointer types (-> 6.11)

## 6.12 Pointer Arithmetic [RVG]

### 6.12.1 Applicability to language

The vulnerabilites described in TR 24772-1 clause 6.12.1 also apply to C++ pointers. Analogous vulnerabilities can also apply to C++ iterators.

Although based on the same implementation principles, iterators provide a layer of abstraction over pointer arithmetic. Their use typically restricts the arithmetic to the safe access to elements of the container. This restriction is enforced by the typical usage, not necessarily by the capability of iterators.

### 6.12.2 Guidance to language users

* Follow the guidance of clause 6.8.2.
* Prefer standard algorithms to hand-written loops
	+ See Core Guideline.xxx
* Use iterators in lieu of pointers and pointer arithmetic. <<<John McF. to provide list of extras.>>>
* Use an iterator that checks against the bounds of the container before performing the intended operation on the container.
* Consider an outright ban on pointer arithmetic due to the error-prone nature of pointer arithmetic.
* Verify that all pointers are assigned a valid memory address for use.

## 6.13 NULL Pointer Dereference [XYH]

### 6.13.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.13 exists in C++,…

C++ provides a number of mechanisms that allow the programmer to create, manipulate and destroy objects without the explicit use of raw pointers.

1. Containers manage memory and separate memory management from the use of objects.
2. The container interface throws an exception if any container cannot be allocated.
3. Smart pointer creation functions allocate heap memory and handle memory management.
4. References provide similar functionality as pointers, but cannot be null.

C++ mechanisms new, by default, throws an exception if the allocated object cannot be created (i.e. if a null pointer would be returned). C++ does provide other allocation mechanism, including C malloc and a non-throwing new, that are not recommended for general use.

See C++ Core Guidelines R: Resource Management, and CERT EXP34-C “Do not dereference null pointers”

### 6.13.2 Guidance to language users

When dereferencing objects of pointer-like types that may contain a null value, follow the guidance from TR 24772-3 clause 6.13.2.

* Avoid the use of direct memory allocation. Prefer the use of library facilities such as std::make\_unique, and std::make\_shared.
* Consider using std::array when the size of the array is known at compile time.
* Consider using std::vector instead of dynamic memory allocation of an array of dynamic size.
* Use references to reduce the number of places where pointers are dereferenced.
* Do not suppress exceptions on memory allocation. If exceptions are suppressed, follow the guidance of TR 24772-3 clause 6.13.2.

## 6.14 Dangling Reference to Heap [XYK]

### 6.14.1 Applicability to language

The vulnerability as expressed in TR 24772-1 and TR 24772-3 C exists in C++. C++, however, provides mechanisms to mitigate the vulnerability.

C++ provides a rich set of types whose objects may dangle, e.g.

* References
* Pointers
* Iterators
* std::string\_view
* gsl::span
* std::reference\_wrapper

We call these types *potentially dangling.*

If the lifetime of a *potentially dangling* *object* ends before its referent’s lifetime ends, then the vulnerability does not apply to that potentially dangling object. This is the primary C++ strategy for avoiding vulnerabilities due to potentially dangling objects. For example, passing a potentially dangling object as a function parameter/argument(?), and the function does not take ownership of the referent (for example by deleting the referent), then the language guarantees that the lifetime of the referent is longer than the lifetime of the parameter. This does not apply to further copies made to longer-lived potentially dangling objects.

Unanticipated aliasing between parameters, global objects, or function results may sometimes lead to this vulnerability. Because it is not practical to test for or document all disallowed aliasing, a restrictive stance is preferred: “All aliasing that is not explicitly allowed by documentation is forbidden.” Allowances for aliasing may be given to some classes of functions by blanket documentation.  In particular, aliasing is expected and allowed in these classes of functions:

* Assignment and compound assignment operators: the right parameter may alias the left parameter. The function result always refers to the left parameter.
* Functions named “swap”: The two parameters to be swapped may refer to the same object.
* Shift operators used for input and output: the result always refers to the left parameter.
* Prefix increment and decrement operators: the result always refers to the parameter.

// Documentation: “v may refer to a portion of s.  The result refers to s.”

std::string\_view& f( std::string& s, std::string\_view v )

  {

   s = v;  // For operator=, aliasing is allowed by blanket documentation.

   return s;  // Returning a result aliased to the parameter is explicitly allowed.

  }

// Documentation of this function does not mention aliasing

void g( std::string& s, std::string\_view v )

  {

                       // If v were to alias s...

   s.clear();   // ...now v would be dangling!

   s = v;          // And this would have undefined behavior.

  }

void h()

  {

   string hello{ “Hello world!” };

   f( hello, hello ); // OK: aliasing is explicitly allowed by f.

   g( hello, hello );  // wrong: g does not document an allowance

                       // for aliasing, so callers must not pass aliased parameters.

  }

Or even as simple as:

std::string\_view bad("a temporary string"s); // "bad" holds a dangling pointer

### 6.14.2 Guidance to language users

In addition to the guidance provided in TR 24772-1 clause 6.14.5:

* Prefer value types, for example std::string instead of const char\*.
* Adopt a style that makes explicit the ownership and lifetime of all resources.
* Limit the scope of potentially dangling objects.
* Document the referents of potentially dangling objects created by or modified by a function if any potentially dangling object outlives the invocation of that function. See the example above.
* Document any allowable aliasing between the referents of function parameters. Absent such documentation, avoid passing aliased parameters. See the example above.
* When allocating an object, adopt a style that all copies of any potentially dangling reference are guaranteed to be cleaned up before the referent’s lifetime ends.

## 6.15 Arithmetic Wrap-around Error [FIF]

### 6.15.1 Applicability to language

C++ shares the vulnerability with C as documented in TR 24772-1 clause 6.15 and TR 24772-3 clause 6,15.1. C++ specifies that signed overflow is undefined behaviour. Unsigned wraparound is well-defined in C++, but it can result in coding mistakes: Example: 4U – 5U yields a large positive value. The smallest signed negative values might not have a positive counterpart (using the same signed integer type). Combined with implicit conversions or promotions for terms in mixed-type expressions, the semantics combine to produce results that can surprise the user.

The mitigations for wrap-around errors in C++ are different.

C++ allows the definition of class types that embed integers together with the operations that provide the wrapping behaviour intended in an efficient way.

Integral promotion – the addition of 2 unsigned chars will promote to (signed) int and then cast back.

### 6.15.2 Guidance to language users

* If you intend to wrap, use an unsigned type that does not promote to int.
* Document where wraparound is expected for a type.
* Consider creating classes that encapsulate integers and that detect or avoid wraparound errors.
* Consider creating classes that explicitly allow wrap-around behaviour.
* Document code that appears convoluted but has been created to avoid wrapping.
* ES103 Don’t overflow (-> 6.15)
* ES104 Don’t underflow (really overflow negatively) (-> 6.15)

References:

Core Guidelines ES.102 “Use signed types for arithmetic”

Core Guidelines ES.103 “Don’t overflow”

Core Guidelines ES.104 “Don’t underflow”

MISRA C++ 5.19.1

## 6.16 Using Shift Operations for Multiplication and Division [PIK]

### 6.16.1 Applicability to language

The issues for C++ are well defined in TR 24772-1 clause 6.16 *Using Shift Operations for Multiplication and Division [PIK].* Also see clause *6.15 Arithmetic Wrap-around Error [FIF]*.

### 6.16.2 Guidance to language users

The guidance for C++ users is well defined in TR 24772-1 clause 6.16 *Using Shift Operations for Multiplication and Division [PIK].* Also see, *6.15 Arithmetic Wrap-around Error [FIF].*

References:

## 6.17 Choice of Clear Names [NAI]

### 6.17.1 Applicability to language

This subclause requires a complete rewrite to have it reflect C++ issues..

C is somewhat susceptible to errors resulting from the use of similarly appearing names. C does require the declaration of variables before they are used. However, C allows scoping so that a variable that is not declared locally may be resolved to some outer block and a human reviewer may not notice that resolution. Variable name length is implementation specific and so one implementation may resolve names to one length whereas another implementation may resolve names to another length resulting in unintended behaviour.

As with the general case, calls to the wrong subprogram or references to the wrong data element (when missed by human review) can result in unintended behaviour.

### 6.17.2 Guidance to language users

This subclause requires a complete rewrite.

* Use names that are clear and non-confusing.
* Use consistency in choosing names.
* Keep the scope of names as small as reasonable.
* Keep names short and concise in order to make the code easier to understand.
* Use longer names for longer-lived objects.
* Choose names that are appropriately rich in meaning for the context.
* When choosing names, keep in mind that code will be reused and combined in ways that the original developers never imagined.
* Do not differentiate names through only a mixture of case or the presence/absence of an underscore character.
* Do not choose names that conflict with (unreserved) keywords or language-defined library names for the language being used, as follows:
	+ Names that begin with double underscore;
	+ Names that begin with a single underscore followed by an uppercase letter;
	+ Contextual keywords such as module, final and override;
	+ In the global namespace, identifiers commencing with std followed by any string of digits;

Follow common conventions for naming macros:

* + Avoid names for macros that are not all uppercase;
	+ Avoid names that are all uppercase not used for macros;
* Avoid differentiating through characters that are commonly confused visually such as ‘O’ and ‘0’, ‘l’ (lower case ‘L’), ‘I’ (capital ‘I’) and ‘1’, ‘S’ and ‘5’, ‘Z’ and ‘2’, and ‘n’ and ‘h’.
* Adopt or develop coding guidelines to define a common coding style and to avoid the above dangerous practices.

## 6.18 Dead Store [WXQ]

### 6.18.1 Applicability to language

The vulnerability as documented in TR 24772-1 clause 6.18 exists in C++.

For Volatile, what do you do to ensure that a write reaches memory?

Initializing part of an array zeros the rest in C++

For the definition of “dead store” in C++, non-trivial destructors constitute “use of an object” .

### 6.18.2 Guidance to language users

* Use compilers and static analysis tools to identify dead stores in the program.
* Declare variables to be accessed by other execution threads that represent values of type T as std::atomic<T>.
* If variables are intended to be accessed by external devices, declare them as volatile.
* If variables are intended to be used to communicate with signal handlers, declare them as volatile sig\_atomic\_t.
* Declare variables as volatile when they are intentional targets of a store whose value does not appear to be used.

## 6.19 Unused Variable [YZS]

### 6.19.1 Applicability to language

The vulnerability as documented in TR 24772-1 clause 6.19 exists in C++.

### 6.19.2 Guidance to language users

* Follow the guidance of TR 24772-1 clause 6.19.5.
* Resolve all compiler warnings for unused variables.

## 6.20 Identifier Name Reuse [YOW]

### 6.20.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.20 exists in C++, except for the second issue of limited identifier length. In C++ all characters in an identifier are significant.

C++ provides the scope resolution operator ‘::’ to access identifier from non-local scopes.

Overloading and specialization of functions is a cornerstone of C++ generic programming. In this context, the reuse of function names is essential. See clause 6.41 for inheritance issues associated with name reuse.

### 6.20.2 Guidance to language users

* Follow TR 24772-1 clause 6.20, with the exclusion of guidance related to truncated identifiers.
* Qualify names to disambiguate potential conflicts between names introduced from different scopes.
* Document argument-dependent lookup usage where name qualification is not desirable.
* Use modern integrated development environments that inform about the declaration of any identifier occurrence.
* Enable compiler diagnostics that inform about the hiding of declarations.
	+ DCL60-CPP. Obey the one-definition rule (6.21)
	+ DCL40-C. Do not create incompatible declarations of the same function or object (6.21)

## 6.21 Namespace Issues [BJL]

### 6.21.1 Applicability to language

The vulnerability described in TR 24772-1is restricted to the following cases:

* Overloading, where clause 6.20 applies;
* Overriding, where clause 6.41 applies.

In all other cases, C++ compilers are required to diagnose an ambiguity.

### Guidance to language users

* Follow the guidance of clauses 6.20.2 and 6.41.2 as applicable.
	+ DCL60-CPP. Obey the one-definition rule (6.21)
	+ DCL40-C. Do not create incompatible declarations of the same function or object (6.21)

## 6.22 Initialization of Variables [LAV]

### 6.22.1 Applicability to language

The vulnerability as described in TR 24772-1 exists in C++.

C++ provides language capabilities to mitigate the effects of uninitialized variables as follows:

See C++ Core Guidelines ES.20 and CERT C++ Coding Guidelines EXP53-CPP

Need a list of references TBD – (AI – J. Daniel Garcia)

Readers should note that ES.20 and EXP53 are complementary. Both point out that you should always initialize before reading, but ES.20 uses the narrow sense of initialize while EXP53 includes assignment.

### 6.22.2 Guidance to language users

* Follow the guidance provided in
	+ C++ Core Guidelines, section Class hierarchies, and Expressions and Statements and
	+ SEI CERT C++ Coding Standard section EXP53-CPP (and possibly more).

## 6.23 Operator Precedence and Associativity [JCW]

### Applicability to language

*The vulnerability as described in TR 24772-1 clause 6.23 is applicable to C++.*

Operator precedence and associativity in C++ are determined by the C++ grammar. There are four operators that cannot be overloaded (user-defined) '::', '.', '.\*', and '?:'. Due to the large number of operators, one is recommended to consult an operator precedence table when needed, e.g., <https://en.cppreference.com/w/cpp/language/operator_precedence>, since the vulnerability as described in TR 24772-1 Clause 6.23 is applicable to C++. For example, in C++, the bitwise logical and shift operators are sometimes incorrectly treated as having the same precedence as arithmetic operations -- they do not -- the bitwise operators have lower precedence. For example, the following (correct) expression subtracts one from x and then checks if the result is zero:

 x - 1 == 0

which is equivalent to (x - 1) == 0, i.e., x - 1 is done first, then that result is compared to zero. Programmers mistakenly thinking the bitwise operations have the same precedence as arithmetic ones might write:

 x & 1 == 0

intending to perform (x & 1) == 0, but precedence rules result in this evaluating x & (1 == 0) instead. (When in doubt, use parenthesis to ensure the proper evaluation of an expression.)

In addition to the aforementioned, C++ also permits operators to be overloaded when used with user-defined types. While it is not possible to change the precedence, associativity, and number of operands of overloaded operators [C++17, Clause 16.5 [over.oper], para. 6], overloaded operators can be executed differently than built-in operators. For example, overloaded operators lose any built-in operator short-circuiting properties and sequence order guarantees. Similarly overloaded operators and their arguments' evaluations behave as normal function calls, differing from built-in operator evaluation. Note: overloaded assignment falls into this category.

C++ overloading of operators can cause significant issues. One hazard is that the overloaded operators does not automatically connect the inverse operator, such as == and !=, < and >=, and > and <=, etc. Unless the declarer declares all relevant operators, unexpected results can occur. In addition, overloaded operators && and || do not have shortcut semantics and thus behave differently. This is problematic since the safety of a right-hand operand of a short circuit operator often depends on the result of the left-hand operand, e.g. an existence test before the value is read.

### 6.23.2 Guidance to language users

* Follow the guidance provided in TR 24772-1 Clause 6.23.5 [JCW].
* Enable all C++ compiler/tool warnings and static analysis tool diagnostics concerning possible issues with precedence and associativity to help avoid and detect mistakes.
* Even if technically unnecessary, use parentheses around operator (sub)expressions that are known to be or felt likely to be sources of error.
* Avoid overloading the following operators:
	+ && ;
	+ || ;
	+ , ;
* Break up complex expressions and use temporary variables to make complex expressions easier to understand and maintain.

## 6.24 Side-effects and Order of Evaluation of Operands [SAM]

### 6.24.1 Applicability to language

The evaluation of an expression includes (i) its value computation and (ii) its side-effects. The value computation is the value returned by the expression, e.g., the valuation of 3 \* 2 + 1 is 7. The side-effect of an expression are

* Write accesses to objects in that expression,
* Reading a volatile object
* Calling a library I/O function, and
* Calling a function that does any of the above.

 For example consider:

 int i = 2;

 int j = i++;

the evaluation of i++ is 2 and the side-effects are the writing of 3 to i and the initialization of j.

Within an expression, one must ensure an object is stored only once to avoid undefined behaviour, e.g.,

 i = i++ + 5; // undefined behaviour (before C++17)

or

 k = i++ + i--; // undefined behaviour in all versions of C++

and expressions modifying objects can only read the object to determine the value to be stored (e.g., ++i requires reading the value), i.e., other accesses are undefined behaviour, e.g.,

 my\_array[i] = i++; // undefined behaviour (before C++17)

Starting with C++17, the evaluation order of an expression involving overloaded operators preserves the sequenced before behaviour of the built-in operator:

my\_array[i] = i++;

 my\_array[i++] = i++;

say I = 10 before the expression

evaluate RHS i++ -- i is 11

evaluate my\_array[i++] --evaluates my\_array[11], then assigns i to 12

my\_array[11] is assigned 10

i.e., assignment is sequenced after the value computation of the right and left operands and before the value computation of the assignment expression; and, the right operand is sequenced before the left operand. [C++17, Clause 8.18 [expr.ass], para. 1] Since this is the built-in operator, this statement can be thought of as:

 Compute value of right-hand-side: i++ (e.g., integer value).

 Compute value of left-hand-side: my\_array[i] (e.g., memory address).

 Apply side-effects of i++.

 Apply side-effects of the assignment.

In general, one should follow commonly-stated C/C++ advice of never reading from and writing to the same object within an expression to avoid potential vulnerabilities. Often breaking the expression into separate statements achieves clear and clean semantics, e.g.,

 ++i;

 my\_array[i] = i;

or

 my\_array[i] = i;

 ++i;

makes it unambiguous what the value of i is during the array assignment and eliminates the possibility of vulnerabilities.

In addition, it is important to note that overloading an operator disables short-circuiting behaviours (e.g., built-in boolean operators): those operators' operands are all evaluated before the operator itself.

The C++ built-in (two-argument) Boolean operators (e.g., && and ||)as well as <type\_traits>'s std::conjunction and std::disjunction operations are all short-circuiting, i.e., if the value of an earlier (from left-to-right) operand of an operation determines the result of the operation, then all remaining arguments are not evaluated.

Typically this allows one to write code like this, e.g.,

 int \*p;

 // ...

 if (p != nullptr && \*p != 0) {

 /\* do something \*/

 }

i.e., if p is nullptr, then \*p != 0 is never executed, thus, avoiding undefined behaviour. Only when p is not nullptr is \*p != 0 is evaluated. It must be stressed that this only applies to the built-in && and || operators: user-defined operator overloads always evaluate all operands first.

Consequently should one want to always evaluate all operands of a boolean expression, one should not write code like this:

 bool x = foo() && bar();

where foo() and bar() are functions that return something convertible to bool. In this expression, if foo() returns false, then bar() will never be executed; --only when foo() returns true will bar() be executed. Similarly for ||:

 bool y = foo() || bar();

i.e., only when foo() returns false will bar() be executed --if foo() returns true then bar() will never be executed. Thus, if both foo() and bar() are both required to be executed, then execute them in separate statements first, e.g.,

 bool foo\_result = foo();

 bool bar\_result = bar();

 bool x = foo\_result && bar\_result;

 bool y = foo\_result || bar\_result;

[Stephen: My write-up here is lengthy but should help get more terse wording... but I note this: C++ operator information is in C++17 Clause 8 and Clause 16.5, ... Also per 16.5.1 para 2. unary and binary forms of the same operator are considered to have the same name so one can hide another from an enclosing scope. Thus, this is also another possible vulnerability.]

C allows expressions to have side effects. If two or more side effects modify the same expression as in:

 int v[10];

 int i;

 /\* … \*/

 i = v[i++];

the behaviour is undefined and this can lead to unexpected results. Either the “i++” is performed first or the assignment i=v[i] is performed first, or some other undefined behaviour occurs. Because the order of evaluation can have drastic effects on the functionality of the code, this can greatly impact portability.

There are several situations in C where the order of evaluation of subexpressions or the order in which side effects take place is unspecified including:

* The order in which the arguments to a function are evaluated (C, Section 6.5.2.2,"Function calls").
* The order of evaluation of the operands in an assignment statement (C, Section 6.5.16,"Assignment operators").
* The order in which any side effects occur among the initialization list expressions is unspecified. In particular, the evaluation order need not be the same as the order of subobject initialization (C, Section 6.7.9, “Initialization").

Because these are unspecified behaviours, testing may give the false impression that the code is working and portable, when it could just be that the values provided cause evaluations to be performed in a particular order that causes side effects to occur as expected.

### 6.24.2 Guidance to language users

* Follow the guidance provided in TR 24772-1 Clause 6.24.5.
* Write expressions so that the same effects will occur under any order of evaluation that C++ permits ~~since~~ side effects can be dependent on an implementation specific order of evaluation.
* Do not overload logical operators (&& and ||), as these overloaded versions will not short-circuit.
* *Ensure that ++, --, @= are used as solitary items in a statement.*
* Be aware to which C++ standard a compiler is compiling code against. This determines the semantics of all operator (overloaded and built-in) expressions.
* Avoid placing multiple operations with side effects in a single expression.
* Write simple code expressions and statements so that within any expression or statement an object is one of:
	+ Read from;
	+ Only modified once; or
	+ Only read and modified by: a single increment; a single decrement; a single simple assignment; or a single compound assignment.
* Be aware that different versions of C++ have different evaluation order specifications and program accordingly.

## 6.25 Likely Incorrect Expression [KOA]

### 6.25.1 Applicability to language

This subclause requires a complete rewrite to have it reflect C++ issues.

C++ has several instances of operators which are similar in structure, but different in meaning. The most common quoted example for C-based languages is the replacement of “==” with “=” in an expression, or confusion between ‘&’ and ‘&&’, ‘|’ and ‘||’, ‘<’, ‘<<’ and ‘<<<’, ‘>’, ‘>>’ and ‘>>>’.

As a general rule, the use of ‘=’, ‘+=’, ‘-=’ in an expression when the operator is not the final assignment to a variable is unsafe since the assignment operator creates side-effects within the expression which are difficult to analyze by a human reader and can be have different results depending upon the order of evaluation of terms within the expression.

C++ provides significant of freedom in constructing statements. This freedom, if misused, can result in unexpected results and potential vulnerabilities.

The flexibility of C++ can obscure the intent of a programmer. Consider:

int x,y;

/\* … \*/

if (x = y){

 /\* … \*/

}

A fair amount of analysis may need to be done to determine whether the programmer intended to do an assignment as part of the if statement (perfectly valid in C++) or whether the programmer made the common mistake of using an “=” instead of a “==”. The major issue with assignment inside ofa term of an expression is that it creates side effects that can cause the expression to evaluate in different orders and create different results on different compilers, or even in different executions with the same implementation.

In order to prevent this confusion, move assignments in contexts that are easily misunderstood outside of Boolean expression. This would change the example code to:

int x,y;

/\* … \*/

x = y;

 if (x == 0) {

 /\* … \*/

 }

This would clearly state what the programmer meant and that the assignment of y to x was intended.

Programmers can easily get in the habit of inserting the “;” statement terminator at the end of statements. However, inadvertently doing this can drastically alter the meaning of code, even though the code is valid as in the following example:

 int a,b;

 /\* … \*/

 if (a == b); // the semi-colon will make this a null statement

 {

 /\* … \*/

 }

Because of the misplaced semi-colon, the code block following the if will always be executed. In this case, it is extremely likely that the programmer did not intend to put the semi-colon there.

Unary ‘+’ on a variable is a no-op, and is possibly a mistype of ‘++’. A unary ‘-‘ on a variable will switch its sign, unless applied to a variable of an unsigned type, in which case WHAT??.

Document with comments any use of ‘+’ or ‘-‘ applied as a unary since (as opposed to the binary ‘+’ or ‘-‘.

* Unary minus on unsigned type (MISRA 5-3-2)
* Size of a pointer

### 6.25.2 Guidance to language users

* Simplify statements with interspersed comments to aid in accurately programming functionality and help future maintainers understand the intent and nuances of the code. The flexibility of C permits a programmer to create extremely complex expressions.
* From Core guidelines:
	+ ES 85 Make empty statements visible
	+ ES 40
	+ ES 41
	+ ES 44 Do not depend on order of evaluation
* Avoid assignments embedded within other statements, as these can be problematic. Each of the following would be clearer and have less potential for problems if the embedded assignments were conducted outside of the expressions:

 int a,b,c,d;

 /\* … \*/

 if ((a == b) || (c = (d-1))) /\* the assignment to c may not

 occur if a is equal to b \*/

 or:

 int a,b,c;

 /\* … \*/

 foo (a=b, c);

 Each may have unintended results.

* Give null statements a source line of their own. This, combined with enforcement by static analysis, would make clearer the intention that the statement was meant to be a null statement.
* Consider the adoption of a coding standard that limits the use of the assignment statement within an expression.

## 6.26 Dead and Deactivated Code [XYQ]

### 6.26.1 Applicability to language

The vulnerability as documented in TR 24772-1 clause 6.26 exists in C++.

### 6.26.2 Guidance to language users

* Follow the guidance of TR 24772-1 clause 6.26.5.

## 6.27 Switch Statements and Static Analysis [CLL]

### 6.27.1 Applicability to language

Because of the way in which the switch-case statement in C++ is structured, it can be relatively easy to unintentionally omit the break statement between cases causing unintended execution of statements for some cases.

The switch statement has the form:

 int abc = someExpression();

 /\* … \*/

 switch (abc) {

 case 1:

 sval = “a”;

 break;

 case 2:

 sval = “b”;

 break;

 case 3:

 sval = “c”;

 break;

 default:

 throw SomeException();

 }

If there isn’t a default case and the switched expression doesn’t match any of the cases, then control simply shifts to the next statement after the switch statement block. Unintentionally omitting a break statement between two cases will cause subsequent cases to be executed until a break or the end of the switch block is reached. This could cause unexpected results.

The attribute [[fallthrough]] expresses the programmer’s intent that the code where it is placed is intended to fall through. If this attribute is not used, compilers typically diagnose the absence of a break statement.

### 6.27.2 Guidance to language users

* Apply the guidance provided in TR 24772-1 clause 6.27.5
* Use [[fallthrough]] wherever fall-through is intended.
* Terminate every case with either a flow control transfer or [[fallthrough]] as illustrated in the following example:

int i;

. . .

switch (i) {

 case 1:

 [[fallthrough]]; // documents the intended fallthrough.

 case 2:

 i++;

 break;

 case 3:

 j++;

 [[fallthrough]]; // documents the intended fallthrough.

 case 4: //other code

 return 42;

 default: throw CaseNotFound();

 }

* Adopt a style that permits your language processor and analysis tools to verify that all cases are covered. Where this is not possible, use a default clause that diagnoses the error.

See also the C++ Core Guidelines ES.78

## 6.28 Demarcation of Control Flow [EOJ]

### 6.28.1 Applicability to language

C++ lacks a keyword to be used as an explicit terminator. Therefore, it may not be readily apparent which statements are part of a loop construct or an if statement.

Consider the following sections of code:

 int foo(int a, const int \*b) {

 int i=0;

 // . . .

 a = 0;

 for (i=0; i<10; i++);

 {

 a = a + b[i];

 }

 int c = 0;

 int x = 0;

 for (int i=0; i<10; i++)

 c = c + b[i];

 x+= c;

 }

At first it may appear that after the first loop, a will be a sum of the numbers b[0]to b[9]. However, even though the code is laid out so that the a = a + b[i] code appears to be within the for loop, the “;” at the end of the for statement causes the loop to be on a null statement (the “;”) and the a = a + b[i];statement to only be executed once. Similarly, the indentation leads us to believe that that assignment to x is part of the second loop, but it is not. These mistakes may be readily apparent during development or testing. More subtle cases may not be as readily apparent leading to unexpected results.

If statements in C are also susceptible to control flow problems since there isn’t a requirement in C for there to be an else statement for every if statement. An else statement in C always belong to the most recent if statement without an else. However, the situation could occur where it is not readily apparent to which if statement an else belongs due to the way the code is indented or aligned.

Similar issues arise for if-statements, particularly during maintenance, for example:

int a,b,i;

// . . .

if (i == 10){

 a = 5;

 b = 10; // added later, but correct since within the {…}

 }

 else

 a = 10;

 b = 5; // added later, intended to be part of the else clause

If the assignments to b were added later and were expected to be part of each if and else clause (they are indented as such), the above code is incorrect: the assignment to b that was intended to be in the else clause is unconditionally executed.

### 6.28.2 Guidance to language users

* Follow the rules provided in TR 24772-1 clause 6.28.5.
* Enclose the bodies of if, else, while, for, and similar in braces. This will reduce confusion and potential problems when modifying the software.
* Declare loop variables in the initializer of the loop statement
* Prefer the standard library algorithms over hand-crafted loops.

See also the C++ Core Guidelines ES.85, ES.71, ES.74, ES.1 and ES.2

## 6.29 Loop Control Variables [TEX]

### 6.29.1 Applicability to language

C++ allows the modification of loop control variables within a loop. Though this is usually not considered good programming practice as it can cause unexpected problems, the flexibility of C++ expects the programmer to use this capability responsibly.

Since the modification of a loop control variable within a loop is infrequently encountered, reviewers of C++ code may not expect it and hence miss noticing the modification. Modifying the loop control variable can cause unexpected results if not carefully done. In C++, the following is valid:

 int a;

 for (int i=1; i<10; i++){

 …

 if (a > 7)

 i = 10;

 …

 }

which would cause the for loop to exit once a is greater than 7 regardless of the number of iterations that have occurred.

C++ also permits the use of multiple variable of the same type in the loop header

Mitigation – range for statement – document with an example (see ES.71) – Gabriel

### 6.29.2 Guidance to language users

* Apply the guidance of TR 24772-1 clause 6.29.5.
* Do not modify a loop control variable within a loop. Even though the capability exists in C, it is still considered to be a poor programming practice.
* Use a range for loop in preference to general loops
* Alternatively, use std library functions copy, accumulate, transform, for\_each, etc. in preference to general loops.
* Something about multiple loop control variables in the same loop?

See also the C++ Core Guidelines ES.71, ES.86,

## 6.30 Off-by-one Error [XZH]

### 6.30.1 Applicability to language

Arrays are a common place for off by one errors to manifest. In C, arrays are indexed starting at 0, causing the common mistake of looping from 0 to the size of the array as in:

 int foo() {

 int a[10];

 int i;

 for (i=0, i<=10, i++)

 …

 return (0);

 }

C++ mitigates the issue of sentinel values in strings document in TR 24772-1 by providing the string class and the string\_view class.

C++ does not flag accesses outside of array bounds, so an off by one error may not be as detectable in C++ as in some other languages. Several good and freely available tools can be used to help detect accesses beyond the bounds of arrays that are caused by an off by one error. However, such tools will not help in the case where only a portion of the array is used and the access is still within the bounds of the array.

C++ mitigates these issues by providing

* Range-based for loops
* Std algorithms
* Iterator style loops terminated by !=
* Container classes
* gsl::span (soon to be std::span)

### 6.30.2 Guidance to language users

* Follow the guidance of TR 24772-1 clause 6.30.5.
* Use careful programming, testing of border conditions, and static analysis tools to detect off by one errors in C++.
* Use range-based for loops, Std algorithms, iterator style loops terminated by !=, or container classes in preference to C-style arrays and structures.

See also the C++ Core guidelines ES.1, ES.42, ES.71, SL.con.3 (more to come)

## 6.31 Structured Programming [EWD]

### 6.31.1 Applicability to language

It is as easy to write structured programs in C++ as it is not to. C++ contains the goto statement, which can create unstructured code. Also, C has continue, break, and return that can create a complicated control flow, when used in an undisciplined manner. Spaghetti code can be more difficult for static analyzers to analyze and is sometimes used on purpose to intentionally obfuscate the functionality of software. Code that has been modified multiple times by an assortment of programmers to add or remove functionality or to fix problems can be prone to become unstructured.

Because unstructured code in can cause problems for analyzers, both automated and human, of code, problems with the code may not be detected as readily or at all as would be the case if the software was written in a structured manner.

### 6.31.2 Guidance to language users

* Write clear and concise structured code to make code as understandable as possible.
* Avoid the use of longjmp
* Avoid the use of goto except in the case of exiting a nested loop.

See also the C++ Core guidelines ES.76, ES.77, SL.C.1

## 6.32 Passing Parameters and Return Values [CSJ]

### 6.32.1 Applicability to language

C++ provides both *call by value* and *call by reference*  parameter passing. The argument is evaluated to initialize the formal parameter (in the first case) or bound to the formal parameter (in the second case) of the function that is being called. A formal parameter behaves like a local variable.

An object can be modified in a function by passing the address to the object to the function, for example

 void swap(int \*x, int \*y) { // C-style

 int t = \*x;

 \*x = \*y;

 \*y = t;

 }

A call to this function is swap( &a, &b);

In a preferred style (below), an object may be passed to a function by reference, which eliminates many of the problems enumerated in TR 24772-1 clause 6.32.1 and 6.32.3.

void swap(int & x, int & y) { // C++-style which is like std::swap

 int t = x;

 x = y;

 y = t;

 }

This function is called by swap(a,b);

Where x and y are integer pointer formal parameters, and \*x and \*y in the swap()function body dereference the pointers to access the integers.

C macros use a *call by name* parameter passing; a call to the macro replaces the macro by the body of the macro. This is called *macro expansion*. Macro expansion is applied to the program source text and amounts to the substitution of the formal parameters with the actual parameter expressions. Formal parameters are often parenthesized to avoid syntax issues after the expansion. Call by name parameter passing reevaluates the actual parameter expression each time the formal parameter is read.

*Paragraph about the violation of the keyword “restrict” in Part 3. – C++ does not have this keyword. Think about the issue.*

### 6.32.2 Guidance to language users

* Follow the advice of TR 24772-1 clause 6.32.5.
* Use caution for reevaluation of function calls in parameters with macros.
* Use caution when passing the address of an object. The object passed could be an alias[[2]](#footnote-2). Aliases can be avoided by following the respective guidelines of TR 24772-1 Clause 6.32.5.

See also the C++ Core Guidelines F.7 through F.48.

## 6.33 Dangling References to Stack Frames [DCM]

### 6.33.1 Applicability to language

C++ allows one variable to refer to another variable. For example, a pointer variable can contain the address of another variable; a reference can be bound to a variable; and an iterator can point to a portion of a variable (in this case a container). Should the referencing variable outlive the referenced variable, the subsequent operations through the referencing variable will have undefined behavior.

For example

int \*bad\_pointer() {
  int a = 0;
  return &a;
 }

int& bad\_reference() {
  int a = 0;
  return a;
 }

std::array<int,3>::iterator bad\_iterator()
 {
  std::array<int,3> a = { 1, 2, 3 };
  return a.begin();
 }

auto bad\_lambda() {

    int x = 0;

    return [&] { x = 1; };

}

void erroneous\_use() {
  std::cout << \*bad\_pointer();
  std::cout << bad\_reference();
  std::cout << \*bad\_iterator();

  std::cout << bad\_lambda()();
 }

### 6.33.2 Guidance to language users

* Do not assign the address of an object, or reference to any entity where the referencing entity persists after the object has ceased to exist. This is done in order to avoid the possibility of a dangling reference.
* Do not return the address of a local variable as the result of a function call.
* Do not return a local variable as the result of a function returning a reference type
* Avoid capturing by reference in lambdas that will be used non-locally, including return, or passing it to another thread, or stored in dynamic memory

See also C++ Core Guidelines F.53, …

## 6.34 Subprogram Signature Mismatch [OTR]

### 6.34.1 Applicability to language

In general, there must be a match between the number of parameters in a function call and the number of arguments in the function definition, with the exception of va\_arg functions f(…).

C++ allows a variable number of arguments in function calls. A good example of a va\_arg function the printf() function. This is specified in the function call by terminating the list of parameters with an ellipsis (, ...). After the comma, no information about the number or types of the parameters is supplied. The use of this feature outside of special situations can be the basis for vulnerabilities.

### 6.34.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.34.5.
* Avoid va\_arg functions .

See also C++ Core Guidelines F.55.

## 6.35 Recursion [GDL]

### 6.35.1 Applicability to language

C++ permits recursion, hence is subject to the problems described in 6.35.

### 6.35.2 Guidance to language users

* Apply the guidance described in TR 24772-1 clause 6.35.5.

## 6.36 Ignored Error Status and Unhandled Exceptions [OYB]

### 6.36.1 Applicability to language

By default, C++ has the C weakness of permitting the call to a function that returns an error code without capturing the return value in a variable. For example

 errnum foo( int a, int b);

. . .

foo(x, y); // failure to capture the return error code.

C++ offers as a mitigating mechanism the [[nodiscard]] attribute. This attribute indicates that the function result must not be discarded.

[[nodiscard]] errnum foo( int a, int b);

. . .

foo(x, y); // compiler error.

if( auto e = foo(a,b); e == 0) { // no compiler error

// success

}

else {

// handle errors

}

 *Should we include a discussion about C++ error\_code??? AI – Michael Wong*

Discuss global error states, such as errno (which is thread-local) but still static.

Global state for error codes is hard to manage and it is easy to forget to check it (C++ Core Guidelines E.28).

C++ offers a set of library-defined exceptions for error conditions that may be detected by checks that are performed by the standard library. In addition, the programmer may define exceptions that are appropriate for their application. These exceptions are handled using an exception handler. Exceptions may be handled in the environment where the exception occurs or may be propagated out to an enclosing scope.

### 6.36.2 Guidance to language users

* Follow the mitigation mechanisms of subclause 6.36.5 of TR 24772-1.
* Check the returned error status upon return from a function. The C standard library functions provide an error status as the return value and sometimes in an additional global error value.
* Use static analysis tools to detect and report missing or ineffective error detection or handling.
* Avoid error handling based on global state.
* Use [[nodiscard]] to prevent callers from ignoring error values.
* Prefer throwing exceptions to returning error values.
* Use destructors to manage the finalization of the current context upon exit, whether erroneous or not.
* Return error values from each enclosing function until an alternative strategy is available.  Consider throwing an exception in lieu of returning an error value.
* Handle exceptions at each function where an alternative strategy is available. In functions where no alternative strategy is available, do not catch the exception.
* Consider termination as a last resort strategy for main or for noexcept functions.
* Notify higher level constructs before a thread is allowed to terminate.
* Consider the use of an exception\_ptr object to transport an exception from the terminating thread to another thread for further processing.

See also C++ Core Guidelines E.1, E.2, E.5, E.6, E.13, E.17, E.19, E.25, and E.28.

## 6.37 Type-breaking Reinterpretation of Data [AMV]

### 6.37.1 Applicability to language

This subclause requires a complete rewrite to have it reflect C++ issues.

The primary way in C that a reinterpretation of data is accomplished is through a union which may be used to interpret the same piece of memory in multiple ways. If the use of the union members is not managed carefully, then unexpected and erroneous results may occur.

C allows the use of pointers to memory so that an integer pointer could be used to manipulate character data. This could lead to a mistake in the logic that is used to interpret the data leading to unexpected and erroneous results.

C Part says

*Wait for Gabriel to help analyze this.*

### 6.37.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.38.5.
* When using unions, implement an explicit discriminant and check its value before accessing the data in the union.
* Ensure through static analysis that arbitrary pointer casts return a type compatible with the source.
* Avoid the use of C-style casts and reinterpret\_cast

## 6.38 Deep vs. Shallow Copying [YAN]

### 6.38.1 Applicability to Language

This vulnerability only arises in C++ when there is a mismatch between the object’s copy semantics and the programmer’s intent. (references to Core Guidelines C.22)

C++ objects, by default, are copied member-wise. Each class type may define its own copy, move and assignment operations, allowing a class author to choose an appropriate depth for these operations. Class member types should be chosen to have copy and move semantics that support the semantics of the enclosing class.

<This may belong elsewhere – TBD> C++ provides the “string view” mechanism as safer pointers to strings. Updates through string view are prohibited, but the initial non “view” value can be updated and this change will be seen by all viewers, even if they are dependent on fixed value.

Note: in C++, this is more commonly known as member-wise copying vs semantic copying, or owning vs observing rights.

Note: Why CERT does not address this issue – involves programmer intent and not readily tool-checkable.

### 6.38.2 Guidance to language users

* Prefer the composition of most types from types that have either value semantics or semantics that support the intended copy and move semantics of the enclosing type.
* When the above is not achievable, ensure that the copy assignment operator, copy constructor, move assignment operator, move constructor and destructor provide the desired semantics.
* Avoid the use of raw pointers with the copy operation and (finish or delete)
* Follow the guidance of C++ core guidelines C.20, C.22, C.32, C.67
* *<This may belong elsewhere – TBD>* Avoid updating the value of a string while there are valid string views in existence.

## 6.39 Memory Leak and Heap Fragmentation [XYL]

### 6.39.1 Applicability to language

C++ uses destructors, and a pattern called Resource Acquisition Is Initialization (RAII) which performs recovery of resources. Destructors (and therefore memory and resource releases) are deterministically ordered with respect to other events on their thread. Object destructors will not be called

* When an unhandled exception escapes its thread of execution
* Under conditions of abnormal termination

See CERT ERR50-CPP for list of cases.

The memory leak vulnerability documented in TR24772-1 clause 6.39 exists in C++, unless the programmer takes steps to avoid it. The steps mentioned above will mitigate most memory leak issues.

The mechanisms std::shared\_ptr and std::shared\_future and similarly constructed reference-counting user code do not detect cycles which will cause leaks because the shared pointers (and hence what they point to) will not be destroyed.

### 6.39.2 Guidance to language users

* Use containers and smart pointers in preference to direct (manual) memory management.
* Follow C++ Core guidelines section R and CERT MEM51.
* For heap fragmentation issues, follow the guidance of TR 24772-1 clause 6.39.5. In particular, create pools of fixed size with user-defined operators new and operators delete.
* Use dynamic analysis tools to detect cycles.
* Break cycles, for example by using std::weak\_ptr or appropriate weak pointers.
* Use std::abort() or std::terminate() and related functions only in extreme situations. See CERT ERR50-CPP for list of cases.
* Use debugging tools such as leak detectors to help identify unreachable memory.

## 6.40 Templates and Generics [SYM]

### 6.40.1 Applicability to language

The vulnerability as described in ISO/IEC TR 24772-1 clause 6.40 exists in C++. C++ provides the facility *Templates* to support the generic programming methodology. C++ Templates can be variadic templates. C++ Templates can accept as arguments types, templates, and values, including addresses of global variables. *Concepts* improve template argument checking at compile time.

Programming language compilers help programmers avoid run-time errors by performing static analysis on the code and generating diagnostics that prevent run-time errors. Therefore, the goal of the C++ language is to help transition code towards compile-time analysis instead of relying on runtime executions that result in run-time failures. Templates are one of the favourable mechanisms to achieve this goal of maximizing compile-time analysis to reduce or eliminate run-time analysis.

When used appropriately, they are suitable for embedded and safety critical systems;

* They provide type safe generic in contrast to legacy void\*-based or macro-based genericity;
* They have no runtime overhead for inline operations; and
* They have no memory used or code generated for unused operations which are both critical in limited resource systems.

Excessive use of templates can lead to cognitive overload in terms of learning, understanding and the maintainability of the code. This clause provides explanation and guidance to mitigate problems that can arise.

From TR 24772-1: “Confusion, and hence potential vulnerability, can arise where the instantiated code is apparently invalid, but does not result in a compiler error. For example, a generic class defines a set of members, a subset of which rely on a particular property of the instantiation type (such as a generic container class with a sort member function, only the sort function relies on the instantiating type having a defined relational operator). In some languages, such as C++, if the generic is instantiated with a type that does not meet all the requirements but the program never subsequently makes use of the subset of members that rely on the property of the instantiating type, the code will compile and execute (for example, the generic container is instantiated with a user defined class that does not define a relational operator, but the program never calls the sort member of this instantiation). When the code is reviewed the generic class will appear to reference a member of the instantiating type that does not exist."

The above paragraph does not correctly characterise the issue. In an implicit instantiation of a class template, only those functions that are ODR (one definition rule)-used are instantiated. Therefore, a template argument need not provide all of the operations used by all possible member functions of the class template. This case is exhibited in the standard library. std::vector<T>::emplace requires an accessible constructor of T. The Factory pattern makes constructors of a class C inaccessible to anyone but the factory functions. Still it is desirable to use std::vector with objects of type C. In this case, emplace cannot be used, which is clearly the intended behaviour of the class because objects of type C should only be constructed by the factory functions.

It is the nature of templates that every method that is not called is not compiled. If the program is changed such that a function is later ODR-used and the program recompiled, Three possible outcomes are:

* The program compiles and executes safely;
* The program fails to compile; or
* The program executes erroneously.

Vulnerabilities that arise from the third case are covered elsewhere in the document(list), however, in the case of templates, the fact that code is written and may not be instantiated for a long time since code that invokes it has not been written, errors may not appear until later.

Features such as concepts and “static\_assert” can be used to mitigate the issue.

*Mitigation – ensure complete coverage of Template code with unit tests*

The third case is not specific to templates, except that the problem may be hidden is addressed in 6.2.

In the above paragraph, cases can arise where the programmer has not provided a type with all of the operations needed to function correctly in the template. For example,

template < typename ForwardIterator , typename Value >

 ForwardIterator find (ForwardIterator first, ForwardIterator last, const Value & val)

 {

 while ( first != last && \* first == val )

 ++ first ;

 return first

 }

The standard says that

* the first template argument must be a forward iterator.
* the second template argument type must be comparable to the value type of that iterator using ==.

These requirements are implicit and all the compiler has to go by is their use in the function

body. The result is great flexibility, splendid generated code for correct calls, and spectacularly bad

error messages for incorrect calls. The obvious solution is to specify the first two requirements as

part of the template’s interface:

template < forward\_iterator Iter , typename Value >

 requires equality\_comparable <Value, Iter :: value\_type >

forward\_iterator find ( Iter first, Iter last, const Value & val );

This is roughly what C++20 offers. Note the equality\_comparable concept. It captures the

required relationship between the two template arguments.

The general C++11 static\_assert can also be used:

 static\_assert ( Equality\_comparable <int >); // succeeds

 static\_assert < Eeqality\_comparable <int ,long >); // succeeds

 struct S { int a; };

 static\_assert ( Equality\_comparable <S>); // fails because structs don't

 // automatically get == & !=

**Discussion of “ranges”**

Using the example above, the first two function arguments must denote a sequence. To express this requirement (that [first:last) is a sequence), requires a library extension. C++20 offers that in the Ranges standard-library component (§9.3.5). Hence in the example below we replace iterators with ranges:

template < range R, typename Value >

requires equality\_comparable <Value , Range :: value\_type >

forward\_iterator find (R r, const Value & val )

{

auto first = begin (r);

 auto last = end (r);

 while ( first != last && \* first == val )

 ++ first ;

 return first

}

**Other issues**

Overload resolution of templates differs from the overload resolution of non-template code. This can result in methods or operators being chosen that are unexpected by the developer.

A template constructor is never a copy or move constructor and hence does not prevent the implicit definition of a copy or move constructor even if the constructor looks similar.

There are cases, however, where the instantiation of a template with a parameter that matches syntactic but not semantic requirements, such as ++ for a pointer, will compile and execute but will be wrong and can result in undefined behaviour. Other examples relate to incomplete class types, see clause 6.2 *Type system.*

Another issue – likely to be placed elsewhere and referenced here

Standard vector construction issue where braces are replaced with curly braces. (Paul)

Guidance – use braces initially and only use other format if the compiler fails ???

*(Examples needed*)

Core guidelines (fill in references)

I.9 T.10, T.11, 12, 13, T.20, T.21, T.22, T.23, T.24, T.25, T.26, T.30, T.31 – forward to Clive.

(*We may wish to summarize)*

### 6.40.2 Guidance to language users

* Use static analysis tools to diagnose the use of inappropriate types in a template interface
* *Cyclic dependencies and ODR (one definition rule) use …*
* Consider using *concepts* for each template type parameter
	+ Create and use concepts that specify “meaningful” semantics
* Write templates that check if a specific template argument fulfills the minimal syntactic requirements for the template, for example by using static\_assert with type\_traits.
* Be aware that a constructor template or assignment operator function template will not replace compiler-provided special member functions
* For generic operator functions, consider providing them as hidden friends through mix-in class templates.

As an example, the following class template defines equality operators for single-element structs (Template parameter U):

**template** <**typename** **U**>

**struct** Eq{

 **friend** **constexpr** b**ool**

 **operator==**(**U** **const** &l, **U** **const**& r) **noexcept** {

 **auto** **const** &[vl]=l;

 **auto** **const** &[vr]=r;

 **return** vl == vr;

 }

 **friend** **constexpr** b**ool**

 **operator!=**(**U** **const** &l, **U** **const**& r) **noexcept** {

 **return** !(l==r);

 }

};

// usage of the above, check that type X obtains operator== and operator!=

**struct** X: Eq<X>{

 **int** val;

 **constexpr** **explicit**

 **X**(**int** v):val{v}{}

};

 **static\_assert**(X{42} != X{43} && X{42} == X{42} );

The advantage in using the above mechanism is that these overloads are only visible comparing objects of type X, and not for other types. Implementing them as free functions increases likelihood that implicit conversions will result in the wrong function being called.

* Use qualified-id or this-> to refer to names that may be found in a dependent base

*Needs an example and explanation in 6.40.1 (AI Paul)*

* For template specialization, declared the specialization:
	1. In the same file as the primary template; or
	2. In the same file as the user-defined type for which the specialization is declared.
* Do not specialize function templates, except when specialization is on a non-deduceable template parameter

## 6.41 Inheritance [RIP]

## 6.41.1 Applicability to language

Inheritance, the ability to create enhanced and/or restricted object classes based on existing object classes, can introduce a number of vulnerabilities, both inadvertent and malicious. Because inheritance allows the overriding of methods of the parent class and because object-oriented systems are designed to separate and encapsulate code and data, it can be difficult to determine where in the hierarchy an invoked method is actually defined.

Also, since an overriding method does not need to call the method in the parent class that has been overridden, essential manipulation of class data may be bypassed.

This can be especially dangerous in copy assignment operator and move assignment operators and in particular when private data components (that is, data components not visible to methods of subclasses) of the parent class are left unchanged. Serious violations of type invariants can arise as a consequence.

Multiple inheritance adds additional complexities to the resolution of method invocations.

The use of inheritance can lead to an exploitable application vulnerability or negatively impact system safety in several ways:

* Execution of malicious redefinitions, which can occur through the insertion of a class into the class hierarchy that overrides commonly called methods in the parent classes.
	+ mitigation – make member functions ‘final’,
	+ reduce the use of inheritance
* Accidental override, where a member function is defined that inadvertently overrides a member function that has already been defined in a parent class.
	+ Mitigation – use “override” and “final” keywords on member functions to generate compiler diagnostics when overriding is accidental
* Accidental failure to override, when a method is incorrectly named or the parameters are not defined properly, and thus does not override a member function in a parent class.
	+ Mitigation – use “override” and “final” keywords on member functions to generate compiler diagnostics when overriding is accidental
* Breaking of class invariants, which can be caused by redefining methods that assign, move, or validate class data without including the assigning, moving or validating in the overriding member function. This applies particularly to class invariants involving data of the parent class not visible in methods of the subclass. Inherited methods of the parent that have access to these “private” components will likely fail, if the components are set inappropriately.
	+ Mitigation – if any class invariant depends upon a value of a data member, then make that member private
* Direct reading and writing of visible class members when matching getting and setting member functions include additional functionality.
	+ Guidance: make data members private and provide a public interface to access them that preserves class invariants.

These vulnerabilities can increase dramatically as the complexity of the hierarchy increases, especially in the use of multiple inheritance.

As member functions are inherited from multiple chains of ancestors, the determination of which member function implementations exist and are being called, becomes increasingly more difficult for the programmer. Understanding which member functions and data members apply to a given (sub)class becomes exceedingly difficult if these methods or components are inherited homographs (i.e., data components with identical names or member functions with identical signatures). Misunderstandings lead to inadvertent coding errors. The complexity increases even more when multiple inheritance is used to model “has-a“ relationships (see subclause [6.42 Violations of the Liskov substitution principle [BLP])](#_6.42_Violations_of_1): member functions never intended to be applicable to instances of a subclass are inherited nevertheless. For example, an instance of class aircraftCarrier may be “turn”ed merely because it obtained its propulsion screw by a “has-a“-inheritance with “turn” being an obviously meaningful method for the class of propulsionScrew. Meanwhile the user has a quite different expectation of what it means to turn an aircraft carrier. The complications increase if the carrier inherits twice from the class propulsionScrew because it has two propulsion screws.

Changes in the execution of methods can be introduced by adding an unrelated but homographic member function (with signatures involving implicitly convertible types) anywhere is the hierarchies of ancestor classes during maintenance of the code. Malicious implementations can thus be added with each release of an object-oriented library and affect the behaviour of previously verified code. (see subclause [6.42 Violations of the Liskov substitution principle [BLP])](#_6.42_Violations_of_1)

* Guidance: Keep inheritance hierarchies short
* Guidance: Qualify the program to invoke member functions in explicit parent classes.
* Mitigation: use the ‘= delete’ construct to prevent a member function from being called due to an inheritance.

## 6.41.2 Guidance to language users

* Follow the guidance of 24772-1 clause 41.5.
* Avoid the use of multiple inheritance whenever possible.
* Avoid access to data components when getting and setting functions are available for them.
* Keep inheritance hierarchies short and narrow
* Prefer non-virtual functions to virtual functions
* Use “override” and “final” keywords on member functions to generate compiler diagnostics when overriding is accidental
* Use the ‘= delete’ construct to prevent a member function from being called due to an inheritance.
* If any class invariant depends upon a value of a data member, then make that member private
* Make data members private and provide a public interface to access them that preserves class invariants
* Provide complete documentation of all encapsulated data, and how each method affects that data for each object in the hierarchy.
* Inherit only from trusted sources, and, whenever possible, check the version of the parent classes during compilation and/or initialization.
* Provide a member function that provides versioning information for each class.
* Prohibit the use of public inheritance for “has-a” relationships. Use composition instead for “has-a”-relationships.
* Delegate assigning and moving of the parent’s data components by calling the corresponding operation of the parent type. You must delegate in particular when the parent has data components not visible to methods of the subclass. Alternatively, prohibit assignment and motion for classes intended to be base types. *(clarify – this has 2 possible meanings)*
* Avoid the creation of base classes that are both virtual and non-virtual in the same hierarchy.

## 6.42 Violations of the Liskov Substitution Principle or the Contract Model [BLP]

## 6.42.1 Applicability to language

This vulnerability applies to C++ . It can be mitigated by a style of programming that uses wrapper functions to check preconditions, calls a virtual function to perform the required functionality and subsequently checks the postconditions before returning. An example is provided below.

class Base  {
  private:
     virtual int function\_to\_override( int x ) = 0;
     // ...

  public:
     int interface\_to\_overridden\_function( int x ) {
           check\_preconditions( x );
           const auto saved = data\_saved\_for\_postcondition( x );
           auto result = function\_to\_override( x );
           check\_postconditions( x, saved, result );
           return result;
         }
     // ...
 };

## 6.42.2 Guidance to language users

* Obey all preconditions and postconditions of each member function, whether they are specified in the language or not.
* Prohibit the strengthening of preconditions (specified or not) by overriding member functions.
* Prohibit the weakening of postconditions (specified or not) by overriding member functions.
* Prohibit the use of public inheritance for “has-a” relationships. Use composition for “has-a”-relationships instead.
* Use static analysis tools that identify misuse of inheritance in the contract model.
* Ensure that all invariants of a derived class are preserved by all public operations on its public base classes. If this cannot be ensured, make the base class private, or avoid inheritance.

See also C++ Core Guidelines C.120, C.121, C.122, C.126, C.127, and C.129 through C.133.

## 6.43 Redispatching [PPH]

## 6.43.1 Applicability to language

The vulnerability as described in ISO/IEC TR 24772-1:2019 clause 6.43 exists in C++ for virtual functions, except for constructors and destructors which are not dispatching. An example of the infinite recursion is:

#include <iostream>

class A {

public:

    virtual void f() { std::cout << "A::f()\n"; }

    virtual void g() { std::cout << "A::g()\n"; A::f(); }  //call to f() will not dispatch.

    virtual void h() { std::cout << "A::h()\n"; }

    virtual void i() { std::cout << "A::i()\n"; h(); }     //call to f() will dispatch,

                                                           //showing the vulnerability

};

class B : public A {

public:

    void f() override { std::cout << "B::f()\n"; g(); }

    void h() override { std::cout << "B::h()\n"; i(); }

};

int main() {

    B b;

    A \* pA = &b;

    pA->f(); // no problem

    std::cout << "---\n";

    pA->h(); // infinite recursion

}

Overriding - Private virtual functions can be overridden - AI – Paul – write up. May be a namespace issues or a Beaujolais issue.

In C++, the call to a member function can be qualified, as shown in the above example, and avoids the vulnerability.

## 6.43.2 Guidance to language users

* At a call site, consider whether virtual dispatch is desired. If not, construct the call using the qualified name.
* Be suspicious of any call from a virtual member function of the derived class to any member function of any of its base classes.

## 6.44 Polymorphic variables [BKK]

## 6.44.1 Applicability to language

This vulnerability as described in ISO/IEC TR 24772-1:2019 applies to C++. In addition to the upcast and downcast issues addressed in that document, this clause also addresses crosscasting, which is unique(?) to C++.

C++ provides language mitigations to help avoid the problems as follows:

Since C++ supports multiple inheritance, up-casting, down-casting, and cross-casting operations can be used to switch to different (pointer/reference) types in the inheritance hierarchy of a specific object, i.e.,

* up-casting is casting an object to an ancestor type in the object's type inheritance hierarchy.
* down-casting is casting an object to a descendent type in the object's type inheritance hierarchy, and,
* cross-casting is casting an object to a sibling/cousin (possibly removed) type in the object's type inheritance hierarchy.
* Unsafe casts, which include C-style casts and reinterpret\_cast, can cast to unrelated arbitrarily structured types. This allows reading and modifying arbitrary memory areas. See subclause [6.11 Pointer Casting and Pointer Type Changes](#_6.11_Pointer_type_1) [HFC] for more details.

Developers should be aware that virtual member functions can be overridden in derived classes, even if they are private.

Given the following:

struct Z { int z; virtual ~Z() { } };

struct Y { int y; virtual ~Y() { } };

struct A : Z { int a; };

struct B : virtual A { int b; };

struct C : virtual A, Y { int c; };

struct D : B, C { int d; };

D d\_inst;

then these examples demonstrate upcasts, downcasts, and crosscasts:

**Upcasts:**

B\* b\_ptr = &d\_inst; // implicit

C& c\_ref = d\_inst; // implicit

Z\* z\_ptr = static\_cast<Z\*>(&d\_inst);

Y\* y\_ptr = dynamic\_cast<Y\*>(&d\_inst);

**Downcasts:**

D& d\_ref = dynamic\_cast<D&>(\*y\_ptr);

D\* d\_ptr = static\_cast<D\*>(b\_ptr);

**Crosscasts:**

C\* c\_ptr = dynamic\_cast<C\*>(b\_ptr);

Y\* y\_ptr2 = dynamic\_cast<Y\*>(b\_ptr);

C\* c\_ptr = static\_cast<C\*> (static\_cast<D\*>(b\_ptr));

and notes the following about such:

Upcasts**:**

* are the only ones that can be performed implicitly
* can also be done with dynamic\_cast or static\_cast

Downcasts

* are explicit;
* can be done safely with dynamic\_cast;
* dynamic\_cast requires appropriate portions of inheritance to be polymorphic (i.e. has virtual members);
* can be done using static\_cast which is unchecked and may be unsafe;

Crosscasts:

* are explicit
* can be done safely with a single call to dynamic\_cast which requires appropriate portions of inheritance to be polymorphic (i.e. has virtual members).
* can often be done with a chain of static\_casts traversing the inheritance hierarchy, which is almost always unsafe.

## 6.44.2 Guidance to language users

* Follow the advice provided in TR 24772-1 clause 6.44.5.
* If an upcast is needed, prefer using implicit conversion, since an explicit upcast adds unnecessary complexity for the reader.
* If a downcast or a crosscast is needed, prefer using dynamic\_cast since it is checked.
* Ensure that all invariants of a derived class are preserved by all public operations on its public base classes. If this cannot be ensured, make the base class private, or avoid inheritance.
* Do not attempt to navigate class hierarchies using C-style casts or reinterpret\_cast.
* For any class that implements a virtual member function, consider marking that member function final in the definition of that class.

NOTE: This forbids any derived class from redefining the implementation and thereby precludes ambiguity, regardless of whether a call is qualified or not.

NOTE: Making instead the class final contradicts C++ Core Guideline C.139, so is not recommended here.

* Consider declaring virtual methods with protected or private visibility to preclude code from outside of the class hierarchy calling any specific implementation directly.

NOTE: This assumes that within the class hierarchy any qualified call is intentional and is the pattern of a non-public virtual interface.

* OOP52-CPP. Do not delete a polymorphic object without a virtual destructor (->6.44)

See also C++ Core Guidelines ES.48, ES.49, C.146, C.147, C.148 and C.153.

## 6.45 Extra Intrinsics [LRM]

This vulnerability does not apply to C++ for the following reasons:

* When adding intrinsics, implementors are required to follow a specific name pattern that users are not allowed to use in definitions. See C++ standard clause 5.10 [Lex.name].

Operations for swap, sin, cos, conversions float <-> double, saturation,

## 6.46 Argument Passing to Library Functions [TRJ]

Libraries that supply objects or functions are in most cases not required to check the validity of parameters passed to them. In those cases where parameter validation is required there might not be adequate parameter validation.

When calling a library, either the calling function or the library may make assumptions about parameters. For example, it may be assumed by a library that a parameter is non-zero so division by that parameter is performed without checking the value. Sometimes some validation is performed by the calling function, but the library may use the parameters in ways that were unanticipated by the calling function resulting in a potential vulnerability. Even when libraries do validate parameters, their response to an invalid parameter is usually undefined and can cause unanticipated results.

###  Applicability to language

This vulnerability applies in particular to C++ libraries which are designed for high efficiency; responsibility for satisfying the preconditions for most functions rests with the caller. When these preconditions are not met, the result will be undefined behaviour. In addition, error conditions are specified by the language for specific functions, such as raising an exception, returning an error code or a known value, such as NaN.

### 6.46.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.47.5.
* Use translation modes provided the implementation to perform addition analysis or checking, such as contracts checks, or instrumentation of executing code.
* Pay attention to the distinction between precondition violation and error conditions in library documentation. The former results in undefined behaviour; the latter results in defined but possibly unwanted behaviour.

## 6.47 Inter-language Calling [DJS]

### 6.47.1 Applicability to language

C++ is a multi-paradigm language with a number of features that do not interface simply with other language systems. It is left to the implementation team the task of converting the results of these paradigms to constructs that can cross an interface for further processing in other languages.

C++ compilers provide an application binary interface (ABI) that delineates areas of interoperability with other languages or other C++ compiler/runtime systems. An ABI includes calling conventions, data layout, error and exception handling and return conventions, name mangling, data model, initialization of memory, and linkage to operating systems and libraries.

C++ compilers implement a C++ language linkage and a C language linkage. It is implementation-defined what other languages the implementation supports. Alternatively, other language systems provide linkages to C systems[[3]](#footnote-3), leaving the developer the task of channeling everything through this common language system.

### 6.47.2 Guidance to language users

* Follow the guidance contained in TR 24772-1 clause 6.47.5
* Use standard layout types for the interoperable interfaces.
	+ Use language linkage facilities that support the languages being used.
	+ EXP56-CPP. Do not call a function with a mismatched language linkage (-> 6.47)
	+ EXP60-CPP. Do not pass a nonstandard-layout type object across (-> 6.47 ?)
	execution boundaries
* Be aware that the static initialization phase and dynamic initialization for every language system are required before the system begins execution
* Be aware that C++ exceptions are not usually compatible with exceptions in other languages.
* Segregate outgoing cross-language interfacing code into functions that present a C++ interface to the C++ code and implements that interface by calling code compatible with the other language system. Similarly implement incoming cross-language interfaces by providing simplified functions that presents a simplified (C or other language) interface and is implemented by calling C++ code with the correct style.
* Separate the interfacing code from the code containing the main functionality

See also the C++ Core Guidelines CPL.3.

*AI 63-6 – group – add the guidance from 6.47.2 Interoperability into the Core Guidelines.*

## 6.48 Dynamically-linked Code and Self-modifying Code [NYY]

### 6.48.1 Applicability to language

The vulnerability as discussed in TR 24772-1 clause 6.48 is applicable to C++.

### Guidance to language users

Follow the guidance of TR 24772-1 clause 6.48.5.

## 6.49 Library Signature [NSQ]

### 6.49.1 Applicability to language

The vulnerability as enumerated in TR 24772-1 applies to C++.

As a mitigation, the C++ ‘extern “C”’ linkage specifier usually provides simple interoperability with libraries using the C application binary interface (ABI).

### 6.49.2 Guidance to language users

From Part 1, 6.49.5

* Follow the guidance of TR 62443-1 clause 6.49.5.
* Follow the advice of clause 6.47.2 as applicable.

##  Unanticipated Exceptions from Library Routines [HJW]

### 6.50.1 Applicability to language

The vulnerability as documented in TR 24772-1 exists for C++. In particular the issue of the failing dynamic initialization of namespace-scope objects exists in C++.

When dynamic initialization of a namespace-scope object fails with an exception, the exception cannot be caught and the program is terminated. Function-scope static objects, in contrast, are initialized the first time execution passes through the declaration.  Using function-scope static objects in preference to dynamic initialization ensures that there is always an enclosing function that could catch the exception.

exception\_prone\_type troubling\_object;
   // An exception from the constructor could cause termination.

// The following function always returns a reference to the same object,
// which is initialized the first time this function is called.

// If initialization fails, it will be retried on the next call.
exception\_prone\_type& safer\_object()
  {
   static exception\_prone\_type the\_safer\_object;
   return the\_safer\_object;
  }

### 6.50.2 Guidance to language users

* Expect functions not marked noexcept to throw exceptions of arbitrary type. Note that all destructors are implicitly noexcept.
* Follow the advice of 6.36.2 for catching and handling exceptions.
* Prefer function-scope static objects to namespace-scope objects for objects needing dynamic initialization.

## 6.51 Pre-processor Directives [NMP]

### 6.51.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.51 applies to C++.

The C++ pre-processor allows the use of macros that are text-replaced before compilation.

Function-like macros look similar to functions but have different semantics. Because the arguments are text-replaced, expressions passed to a function-like macro may be evaluated multiple times. This can result in unintended and undefined behaviour if the arguments have side effects or are pre-processor directives. Additionally, the arguments and body of function-like macros should be fully parenthesized to avoid unintended and undefined behaviour.

The following code example demonstrates undefined behaviour when a function-like macro is called with arguments that have side-effects (in this case, the increment operator) .

#define CUBE(X) ((X) \* (X) \* (X))

// ...

int i = 2;

int a = 81 / CUBE(++i);

The above example could expand to:

 int a = 81 / ((++i) \* (++i) \* (++i));

which has undefined behaviour so this macro expansion is difficult to predict.

Another mechanism of failure can occur when the arguments within the body of a function-like macro are not fully parenthesized. The following example shows the CUBE macro without parenthesized arguments.

#define CUBE(X) (X \* X \* X)

// ...

int a = CUBE(2 + 1);

This example expands to:

 int a = (2 + 1 \* 2 + 1 \* 2 + 1)

which evaluates to 7 instead of the intended 27.

### 6.51.2 Guidance to language users

* Replace macro-like functions with inline functions where possible.
* Ensure that if a function-like macro must be used, that its arguments and body are parenthesized.
* In a function-like macro, ensure that each argument is evaluated at most once.
* Do not embed pre-processor directives or side-effects such as an assignment, increment/decrement, volatile access, or function call in a function-like macro.

## 6.52 Suppression of Language-defined Run-time Checking [MXB]

### 6.52.1 Applicability to language

With the exception of the macro assert, the vulnerability as described in TR 24772-1 does not apply to C++, because there is no language-defined runtime checking. Macro assert is defined by the standard but is invoked by the programmer, hence is not a language-defined check.

C++ libraries, however, often provide run-time checks which meet the criteria of this vulnerability. Also compilers and other tools commonly provide means to perform such runtime checks.

### 6.51.2 Guidance to language users

Follow the advice provided in TR 24772-1 clause 6.52.5 with respect to library and compiler-provided checks. Note that this will almost always require explicitly enabling the checks.

## 6.53 Provision of Inherently Unsafe Operations [SKL]

### 6.53.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.53 applies to C++. In particular, anything described by ISO/IEC 14882:2017 as “undefined behaviour” is unsafe.

### 6.53.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.53.5.
* Enable checks that warn about unsafe operations.
* Use static analysis tools to detect unsafe constructs.

## 6.54 Obscure Language Features [BRS]

### 6.54.1 Applicability of language

The vulnerability as described in TR 24772-1 clause 6.54 applies to C++.

### 6.54.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.54.5.

## 6.55 Unspecified Behaviour [BQF]

### 6.55.1 Applicability of language

 The vulnerability as described in TR 24772-1 clause 6.55 applies to C++.

### 6.55.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.55.5.

## 6.56 Undefined Behaviour [EWF]

### 6.56.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.56 applies to C++. In ISO/IEC 14882:2017, the terms “undefined behaviour” and “ill-formed, no diagnostic required” expose situations to be avoided.

### 6.56.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.56.5.
* Augment static analysis tool usage with runtime tools such as ASAN (address sanitizer) and related tools.

## 6.57 Implementation–defined Behaviour [FAB]

### 6.57.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.57 applies to C++. In ISO/IEC 14882:2017, the term “implementation-defined” is used to describe implementation-defined behaviour. In addition, the C++ standard provides an index titled “Index of implementation-defined behaviour”.

### 6.57.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.57.5.
* Eliminate to the extent possible any reliance on implementation-defined behaviour from programs in order to increase portability. Even programs that are specifically intended for a particular implementation may in the future be ported to another environment or sections reused for future implementations.

## 6.58 Deprecated Language Features [MEM]

### 6.58.1 Applicability to language

The vulnerability as described in TR 24772-1 clause 6.58 applies to C++. Appendix D “Compatibility features” of ISO/IEC 14882:2017 enumerates the deprecated features. The C++ attribute [[deprecated]] allows library writers and users to mark deprecated declarations.

Although backward compatibility is sometimes offered as an option for compilers so one can avoid changes to code to be compliant with current language specifications, updating the legacy software to the current standard is a better option.

### 6.58.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.58.5.
* Enable compiler options that identify the use of deprecated features.
* Apply the [[deprecated (“*reason*”)]] attribute to obsolete declarations that exist only for backward compatibility.

## 6.59 Concurrency – Activation [CGA]

### 6.59.1 Applicability to language

Michael’s document on concurrency issues in C++ - https://docs.google.com/document/d/14E0BYqsH\_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit

C++ permits concurrent execution through the creation of user-defined threads, hence the vulnerabilities defined by TR 24772-1 apply to C++.

TR 24772-1 uses the term “activation”, which is not a C++ term. We will use the term “creating thread”, and “created thread”.

Recommendation to use C++ concurrency instead of OS concurrency such as processes.

What about OpenMP? – lives on top of OS. Has a decoupled fork-join model where “threads” work on individual work units and deliver their portion at the end of a parallel block.

These clauses will not discuss non-C++ concurrency approaches.

AI – Steve – include a comparison of concurrency approaches in clause 4.

Major differences between Tasks and Threads

* Threads need to be explicitly joined and cleaned up.
* Tasks are owned by a master.
* Tasks and threads share the same scheduler.

C++ uses the fork-join model for task creation. When a thread object is created, either all resources needed to execute the thread have been acquired, or an exception is be thrown in the creating thread. Non-standard threading packages may have the vulnerability that the creating thread does not know if the created thread successfully begins execution.

As soon as the created thread is successfully created, it begins execution and the creating thread resumes execution. If multiple threads are to be created, say in a loop, then each creation is an individual step. Should a resource allocation error occur during the creation of multiple threads, the exception handler associated with the creation will only know which creation cause the exception if the application makes a record of successful creations.

C++ provides the std::thread::get\_id() call to acquire the identity of the created thread, which can be then recorded to make further queries about the state of each active thread.

(The vulnerability documented in TR 24772-1 does not apply in C++ when std::thread is used for threading.)

The second vulnerability of attempts to communicate with non-existing threads does not exist since the std::thread::get\_id() associated with the thread object is unavailable until after creation has completed.

Forking new thread will raise exception in the forking thread if unable to create thread.

There is a scenario where you can think that you are stating a thread, but thread is not launched. Caused by syntax confusion. Declaration of a function that resembles a thread-launching statement (need example) std::thread mythread(backgroundtask()); // declares a function mythread but does not start a thread. Fix std:thread mythread((backgroundtask())); or std::thread mythread{backgroundtask()};

Use of a lambda expression avoids this problem.

std::thread mythread{[](){backgroundtask();}}; // explain // items in [] will be copied and the lifetime will be the same as the background thread.

Affects lifetime. In standard thread creation, a reference to the object that it will work with is made, but both the initiator and the thread must synchronize how they access the object. When a lambda is used, a local copy is made of all objects mentioned in the braces, so a referenced object will be copied.

There is no way to know that the thread is still running from the thread object.

Parallel algorithms rely on “parallel” libraries.

The std::thread clss provides the following methods to query or manipulate a thread:

(constructor) - Construct thread (public member function )

(destructor) - Thread destructor (public member function )

[operator=](http://www.cplusplus.com/reference/thread/thread/operator%3D/) - Move-assign thread (public member function )

get\_id - Get thread id (public member function )

joinable - Check if joinable (public member function ) (Boolean)

join - Join thread (public member function )

detach - Detach thread (public member function )

swap - Swap threads (public member function )

native\_handle - Get the native handle (public member function )

hardware\_concurrency - Detect hardware concurrency (public static member function )

The following comparisons are available to compare two threads:

== (equality),

!= (not equal),

< (less than),

<= (less than or equal),

> (greater), and

>= (greater or equal)

No scoping control, so need to ensure that data passed to another thread could silently go out of scope.

AI – Michael - Verify that the discussion of joinable is complete.

Meta Issue: possibly refer to MISRA C++, etc. for guidance on domain-specific issues and more safe programs.

STL is not thread safe

Can initiate async objects, which can return an exception or a value through a future. Not a full “RAII” class.

There are futures or shared futures.

Package tasks can be detached and return a future or

### 6.59.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.60.5.
* Either join or detach a created thread.
* Use lambda and explicit copy
* When using an async call, delay a local’s destruction until
* Create and manage threads
* Use J-threads (from C++ 20)

## 6.60 Concurrency – Directed termination [CGT]

### 6.60.1 Applicability to language

What about tasks? Explain std::function A functor is a class that can be called like a throw-away parallel object. Can also use in lambda expressions. The creating thread can only retrieve results by using a shared variable or using explicit thread calls to

A thread in C++ runs until completion, either a normal completion or as the result of an unhandled exception. There is no mechanism in the language to terminate another thread. If the threading model is POSIX or some other underlying paradigm, the underlying threading service calls can be used to terminate a thread.

C++ threads use a fork-join model. This means that the initiating thread will wait for the completion of the initiated thread at the join place at the end of the scope that created the initiated thread if a call to join the terminating thread has been made; otherwise the initiating task will have no indication of when the created thread completes.

Programmed mechanisms can be constructed to cause another thread to complete or to raise an exception, such as setting a shared variable to a known state that the target thread reads and then terminates itself.

Are there any language-defined ways to check on the progress of a thread or know if it is executable? Can always use underlying mechanism where available. Future mechanism provides for communication with the spawned thread to know if it terminated returning a value or returning an exception.

## 6.60.2 Guidance to language users

Follow the guidance of 24772-1 Clause 6.59.5

## 6.61 Concurrent Data Access [CGX]

### 6.61.1 Applicability to language

C++ has threading and shared access to variables which have the vulnerabilities described in TR 2772-1 clause 6.61.1. C++ provides features such as atomic (type template) that guarantee the internal consistency of the data and to prevent corruption of data due to potentially interleaved updates to data elements.

What about concurrent data access between tasks?

Programmers should be aware that conversions or manipulations of data items are not always atomic, such as the conversion

*Need the C++ definition of atomic (indivisible access and memory ordering)*

 *and volatile.*

The C++ *atomic* capability can be applied to any basic data type equivalent to char, short, int, long, and long long. When the C++ std::atomic facilities are used, the language guarantees that simultaneous updates and reads to an atomic element will be well-behaved. Atomic does not guarantee the order in which competing reads and/or updates will occur. In order to manage order of access, synchronized locks are required. In order to use the atomic capabilities, each variable must be declared to be of one of the std::atomic types, and the member functions used to compare, load, store or exchange values in an atomic variable.

A volatile qualifier on a variable is used to indicate that updates to the variable may happen at any time and outside of program control, hence two subsequent reads on such a variable may return different results.

Programmers should be aware that even simple data accesses on modern architectures can involve instruction reordering, cache issues, and data alignment issues, hence the acquisition time and order are highly nondeterministic, especially when being accessed by concurrent threads. Any data structure that can be shared between threads should be shown to be accessed by at most one thread at a time or should be protected by synchronization mechanisms such as locks (see 6.63) or atomicity.

Most concurrent programming algorithms require some level of synchronization between threads or tasks when exchanging information, synchronization that “atomic” does not provide. Mechanisms such as monitors, mailboxes, or mutexes (lock with a queue), futures, condition variables, and locks control scheduling of threads or tasks to control order-of-access and to enforce higher levels of cooperation between schedulable entities.

Atomic tied to memory orders.

Mutexes provide mutual exclusion and guaranteed visibility (consistency) of the shared data.

Mutex is a lock-and-release that is usually hidden.

Encapsulate mutexes and data

Thread-level storage (official term thread\_local) has lifetime of the thread. Can exist at local scope or global scope.

For massively parallel concurrency – concurrent access mechanisms not applicable.

No resource management

Exception and exception handling (has some impact on threading)

Memory management issues more complex under concurrency

Volatile should be used for signal handlers to prevent the optimization of replicated accesses to volatile memory. (other) and does not guarantee that the object value will be available to other threads.

Controlling access to shared data (protected or including

Use of volatile (keyword type qualifier) for signal handlers (communicating with hardware?). Prefer volatile for communicating with hardware?

### 6.61.2 Guidance to language users

Much of the guidance is in 6.63 Protocol lock errors

* Follow the guidance of ISO/IEC TR 24772-1 clause 6.62.5.
* Use mutexes, condition variables (convar) in preference to atomic variables to protect data from simultaneous access.
* Use volatile only for synchronizing abstract machine state from memory.
* Avoid the use of volatile for synchronizing access to shared memory between threads (which needs mutex, condvar, or atomics)
	+ See C++ Core guidelines CP.8, CP.200, CP.111,

Avoid relaxed atomic operations whenever possible. Prefer the sequentially consistent std::memory\_order\_seq\_cst instead .

* Use mutexes appropriately to protect accesses to non-atomic shared objects.

## 6.62 Concurrency – Premature Termination [CGS]

### 6.62.1 Applicability to language

A thread will terminate when it completes its assigned method, or when it raises an exception, or when it has been explicitly terminated (how is this done)

Joining a thread causes the joining thread to await the joined thread’s termination before continue. Useful for executing in parallel and then proceeding after the dispatched work is complete, but does not notify the joining task if the termination was premature.

In C++ 2020, methods are provided to instruct one or more threads to terminate. This is not premature termination since the requested thread terminates itself.

C++ 2020 provides callbacks in the form of stop\_callback to notify the setting thread when a thread of interest has been terminated. It also provides stop\_token for a thread to query it is being instructed to terminate.

Any thread can re-throw an exception to be caught by the creator of the terminating thread, (but the parent may have terminated first).

The semantics of C++ is that all children of the main program will terminate if the main program terminates. It is necessary to join the main program to all its children to ensure that children are not silently terminated prematurely.

### 6.62.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.63.5.
* Use low-level operating system primitives or other APIs where available to check that a required thread is still active.
* Make the main program join all created threads that need to complete normally.

## 6.63 Protocol Lock Errors [CGM]

### 6.63.1 Applicability to language

The vulnerability as described in ISO/IEC TR 24772-1 clause 6.63 is applicable to C++.

This subclause requires a complete rewrite to have it reflect C++ issues.

Difference between threads and tasks. Can threads and tasks coexist?

Deadlock with single mutex,

The C standard does not provide hidden protocols. Although the vulnerability does not apply to the C language, there could exist an application vulnerability if a program uses synchronization mechanisms incorrectly. For example:

atomic int a;

int b;

/\* . . . \*/

a += b; // This operation is an atomic read-modify-write of the variable ‘a’.

a = a + b; // This statement contains two accesses to ‘a’ and is not atomic.

### 6.63.2 Guidance to language users

* Follow the guidelines of TR 24772-1 clause 6.63.5.
* Be aware of the operation of each synchronization mechanism, such as the cases where accesses to atomic variables may occur more than once in a statement.
* Use higher level building blocks (such as TBB) in preference to …
* Use the C++ Task mechanism in preference to threads …
* Always put the acquisition and release of mutexes and the data access in a wrapper function. (i.e. *Do not call member functions of std::mutex, std::timed\_mutex, std::recursive\_mutex, std::recursive\_timed\_mutex, std::shared\_mutex and std::shared\_timed\_mutex objects directly.)*
* *Use std::lock(), std::try\_lock() or std::scoped\_lock to acquire multiple mutexes in same scope. (std::lock() permits multiple mutexes at the same time).*
* Use std::lock() only where multiple locks must be locked together and use std::lock\_guard with the std::adopt\_lock argument for all mutexes (needs example) see std::lock() example on cppreference.com.
* *Wrap mutex locks std::lock or std::try\_lock with std****::****lock\_guard, std::unique\_lock or std::shared\_lock with adopt\_lock tag within the same scope*
* If explicit locking are used, ensure that the lock is released on every exit path, including exceptions. Use lock\_guard, scope\_lock and unique\_lock in preference to lock(), unlock(), and try\_lock(),
* ***Do not use platform specific multi-threading facilities***
* ***A thread shall not access objects whose lifetime has expired***
* [0.4.4 [12] Do not destroy objects of the following types std::mutex, std::timed\_mutex, std::recursive\_mutex, std::recursive\_timed\_mutex, std::shared\_mutex, std::shared\_timed\_mutex if object is in locked or shared locked state Do not destroy a mutex while it is locked](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.kj1vcao94oy1)
* [0.4.5 [13] Mutexes locked with std::lock or std::try\_lock shall be wrapped with std::lock\_guard, std::unique\_lock or std::shared\_lock with adopt\_lock tag within the same scope Ensure actively held locks are released on exceptional conditions 12](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.1kfv9jdgd8ib)
* [0.4.6 [14] Do not call virtual functions and callable objects passed by argument of the function within the scope of locked mutex Never call unknown code while holding a lock (e.g., a callback) 12](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.729gvquxakq8)
* [0.4.7 [15] Avoid deadlock by locking in a predefined order](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.920crsa3sscx)
* [0.4.8 [16] Objects of std::lock\_guards, std::unique\_locks, std::shared\_lock and std::scoped\_lock classes shall always be named Remember to name your lock\_guards and unique\_locks](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.xufkh9tsiuk8)
* [0.4.9 [17] Define a mutex together with the data it guards. Use synchronized\_value<T> where possible 13](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.d6qklu5mi3fn)
* [0.4.10 [18] Do not speculatively lock a non-recursive mutex that is already owned by the calling thread 14](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.isn3bwogdp97)
* [0.4.11 [19] There shall be no code path which results in locking of the non-recursive mutex within the scope when this mutex is already locked Within the scope of a lock, ensure that no static path results in a lock of the same mutex 14](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.einwxosnxvhf)
* [0.4.12 [20] The order of nested locks unlock shall form a DAG Ensure that order of nesting of locks in a project forms a DAG 15](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.vq776pxv4nn5)
* [0.4.13 [21] std::recursive\_mutex and std::recursive\_timed\_mutex should not be used Do not use std::recursive\_mutex 17](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.7pf8azwmy8l4)
* [0.4.14 [22] There should be a code path, where at least one member functions is called for std::unique\_lock objects  Only use std::unique\_lock when std::lock\_guard cannot be used 18](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.z9k66jbl65u6)
* [0.5 Conditional variable **19**](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.3icrgfn0r3ae)
* [0.5.1 [23] std::condition\_variable::wait, std::condition\_variable::wait\_for, std::condition\_variable::wait\_until shall always be called with a condition predicate 19](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.g0cdm0bk7l9l)
* [0.5.2 [24] Wrap functions that can spuriously wake up in a loop 20](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.ul4y0p7eoih1)
* [0.5.3 [25] std::conditional\_variable::notify\_one() can be used if all threads must perform the same set of operations after waking up Preserve thread safety and liveness when using condition variables 21](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.13kr181fh926)
* [0.5.4 [26] Do not use std::condition\_variable\_any on a std::mutex](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.yt0hxah53p9e)
* [0.10.1 [35] Source CCG Rule CP.100: Don't use lock-free programming unless you absolutely have to 27](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.3hq5f8vdw7d)
* 0.10.2 [36] Source CCG Rule CP.101: Distrust your hardware/compiler combination
* 0.10.3 [37] Source CCG Rule CP.102: Carefully study the literature
* [0.10.4 [38] Source CCG Rule CP.110: Do not write your own double-checked locking for initialization](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.ca8el9ehijae)
* [0.10.5 [39] Source HIC Rule 18.2.4 Use std::call\_once to ensure a function is called exactly once (rather than the Double-Checked Locking pattern)](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.9syx86lze887)
* [0.10.6 [40] Source CCG Rule CP.111: Use a conventional pattern if you really need double-checked locking](https://docs.google.com/document/d/14E0BYqsH_d7fMKvXvaZWoNWtIC65cYBw0aZp4dlev0Q/edit#heading=h.gg4jn12ygzi5)

## 6.64 Uncontrolled Format String [SHL]

### 6.64.1 Applicability to language

C++ inherits the C libraries which provide a large family of input and output functions that use a control string to interpret the data read or format the output. These strings include all the feature described in TR 24772-1 clause 6.64.1.

C++ provides type-safe alternatives for input/output, which do not use format strings and which should be used in preference, such as
std::cout << “Hello world this is ” << myName << “ and I am “ << myAge << “ years old.\n”;

In addition operator overloading for output operators allows to extend formatting abilities to user-defined types.

AI – 66-2 Peter – Provide alternate examples. I forgot, what I was supposed to provide? Looks good for me for this item. There are vulnerabilities wrt overloading input and output operators, but those do not relate to the vulnerability here.

### 6.64.2 Guidance to language users

Do not use the C libraries from <cstdio>.

# 6.65 Non-constant constants

### 6.65.1 Applicability to language

The vulnerability as documented in ISO/IEC TR 24772-1:2019 clause 8.2 exists in C++.

(Document)

### 6.65.2 Guidance to language users

# 7. Language specific vulnerabilities for C

7.2 Copy/move semantics from Classes. (Peter Sommerlad’s paper at http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2019/p1412r0.pdf

# 8. Implications for standardization

Future standardization efforts should consider:

* Moving in the direction over time to being a more strongly typed language. Much of the use of weak typing is simply convenience to the developer in not having to fully consider the types and uses of variables. Stronger typing forces good programming discipline and clarity about variables while at the same time removing many unexpected run time errors due to implicit conversions. This is not to say that C should be strictly a strongly typed language – some advantages of C are due to the flexibility that weaker typing provides. It is suggested that when enforcement of strong typing does not detract from the good flexibility that C offers (for example, adding an integer to a character to step through a sequence of characters) and is only a convenience for programmers (for example, adding an integer to a floating-point number), then the standard should specify the stronger typed solution.
* A common warning in Annex I should be added for floating-point expressions being used in a Boolean test for equality.
* Modifying or deprecating many of the C standard library functions that make assumptions about the occurrence of a string termination character.
* Define a string construct that does not rely on the null termination character.
* Defining an array type that does automatic bounds checking.
* Deprecating less safe functions such as strcpy() and strcat() where a more secure alternative is available.
* Defining safer and more secure replacement functions such as memncpy() and memncmp() to complement the memcpy() and memcmp() functions (see *6.11.6 Implications for standardization*)
* Defining an array type that does automatic bounds checking.
* Defining functions that contain an extra parameter in memcpy() and memmove() for the maximum number of bytes to copy. In the past, some have suggested that the size of the destination buffer be used as an additional parameter. Some critics state that this solution is easy to circumvent by simply repeating the parameter that was used for the number of bytes to copy as the parameter for the size of the destination buffer. This analysis and criticism is correct. What is needed is a failsafe check as to the maximum number of bytes to copy. There are several reasons for creating new functions with an additional parameter. This would make it easier for static analysis to eliminate those cases where the memory copy could not be a problem (such as when the maximum number of bytes is demonstrably less than the capacity of the receiving buffer). Manual analysis or more involved static analysis could then be used for the remaining situations where the size of the destination buffer may not be sufficient for the maximum number of bytes to copy. This extra parameter may also help in determining which copies could take place among objects that overlap. Such copying is undefined according to the C standard. It is suggested that safer versions of functions that include a restriction max\_n on the number of bytes n to copy (for example, void \*memncpy(void \* restrict s1,const void \* restrict s2,size\_t n), const size\_t max\_n) be added to the standard in addition to retaining the current corresponding functions (for example, memcpy(void \* restrict s1,const void \* restrict s2,size\_t n))). The additional parameter would be consistent with the copying function pairs that have already been created such as strcpy()/strncpy() and strcat()/strncat(). This would allow a safer version of memory copying functions for those applications that want to use them in to facilitate both safer and more secure code and more efficient and accurate static code reviews[[4]](#footnote-4).
* Restrictions on pointer arithmetic that could eliminate common pitfalls. Pointer arithmetic is error-prone and the flexibility that it offers is useful, but some of the flexibility is simply a shortcut that if restricted could lessen the chance of a pointer arithmetic based error.
* Defining a standard way of declaring an attribute to indicate that a variable is intentionally unused.
* A common warning in Annex I should be added for variables with the same name in nested scopes.
* Creating a few standardized precedence orders. Standardizing on a few precedence orders will help to eliminate the confusing intricacies that exist between languages. This would not affect current languages as altering precedence orders in existing languages is too onerous. However, this would set a basis for the future as new languages are created and adopted. Stating that a language uses “ISO precedence order A” would be useful rather than having to spell out the entire precedence order that differs in a conceptually minor way from some other languages, but in a major way when programmers attempt to switch between languages.
* Deprecating the goto statement. The use of the goto construct is often spotlighted as the antithesis of good structured programming. Though its deprecation will not instantly make all C code structured, deprecating the goto and leaving in place the restricted goto variations (for example, break and continue) and possibly adding other restricted goto’s could assist in encouraging safer and more secure C programming in general.
* Defining a “fallthru” construct that will explicitly bind multiple switch cases together and eliminate the need for the break statement. The default would be for a case to break instead of falling through to the next case. Granted this is a major shift in concept, but if it could be accomplished, less unintentional errors would occur.
* Defining an identifier type for loop control that cannot be modified by anything other than the loop control construct would be a relatively minor addition to C that could make C code safer and encourage better structured programming.
* Defining a standardized interface package for interfacing C with many of the top programming languages and a reciprocal package should be developed of the other top languages to interface with C.
* Joining with other languages in developing a standardized set of mechanisms for detecting and treating error conditions so that all languages to the extent possible could use them. Note that this does not mean that all languages should use the same mechanisms as there should be a variety ( label parameters, auxiliary status variables), but each of the mechanisms should be standardized.
* Since fault handling and exiting of a program is common to all languages, it is suggested that common terminology such as the meaning of fail safe, fail hard, fail soft, and so on along with a core API set such as exit, abort, and so on be standardized and coordinated with other languages.
* Deprecating unions. The primary reason for the use of unions to save memory has been diminished considerably as memory has become cheaper and more available. Unions are not statically type safe and are historically known to be a common source of errors, leading to many C programming guidelines specifically prohibiting the use of unions.
* Creating a recognizable naming standard for routines such that one version of a library does parameter checking to the extent possible and another version does no parameter checking. The first version would be considered safer and more secure and the second could be used in certain situations where performance is critical and the checking is assumed to be done in the calling routine. A naming standard could be made such that the library that does parameter checking could be named as usual, say “library\_xyz” and an equivalent version that does not do checking could have a “\_p” appended, such as “library\_xyz\_p”. Without a naming standard such as this, a considerable number of wasted cycles will be conducted doing a double check of parameters or even worse, no checking will be done in both the calling and receiving routines as each is assuming the other is doing the checking.
* Creating an Annex that lists deprecated features.

# Bibliography

[1] ISO/IEC Directives, Part 2, *Rules for the structure and drafting of International Standards*, 2004

[2] ISO/IEC TR 10000‑1, *Information technology — Framework and taxonomy of International Standardized Profiles — Part 1: General principles and documentation framework*

[3] ISO 10241 (all parts), *International terminology standards*

[4] ISO/IEC 9899:2011, *Information technology* — *Programming languages —* C

[5] ISO/IEC 9899:2011/Cor.1:2012, *Technical Corrigendum 1*

[6] ISO/IEC 30170:2012, *Information technology* — *Programming languages —* Ruby

[7] ISO/IEC/IEEE 60559:2011, *Information technology – Microprocessor Systems – Floating-Point arithmetic*

[8] ISO/IEC 1539-1:2010, *Information technology — Programming languages —* Fortran — Part 1: Base language

[9] ISO/IEC 8652:1995, *Information technology — Programming languages —* Ada

[10] ISO/IEC 14882:2011, *Information technology* — *Programming languages —* C++

[11] R. Seacord, *The CERT C Secure Coding Standard*. Boston,MA: Addison-Westley, 2008.

[12] Motor Industry Software Reliability Association. *Guidelines for the Use of the C Language in Vehicle Based Software*, 2012 (third edition)16F[[5]](#footnote-5).

[13] ISO/IEC TR24731–1, *Information technology — Programming languages, their environments and system software interfaces — Extensions to the C library — Part 1: Bounds-checking interfaces*

[14] ISO/IEC TR 15942:2000, *Information technology — Programming languages — Guide for the use of the Ada programming language in high integrity systems*

[15] Joint Strike Fighter Air Vehicle: C++ Coding Standards for the System Development and Demonstration Program. Lockheed Martin Corporation. December 2005.

[16] Motor Industry Software Reliability Association. *Guidelines for the Use of the C++ Language in critical systems*, June 2008

[17] ISO/IEC TR 24718: 2005, *Information technology — Programming languages — Guide for the use of the Ada Ravenscar Profile in high integrity systems*

[18] L. Hatton, Safer C: developing software for high-integrity and safety-critical systems. McGraw-Hill 1995

[19] ISO/IEC 15291:1999, *Information technology — Programming languages — Ada Semantic Interface Specification (ASIS)*

[20] Software Considerations in Airborne Systems and Equipment Certification. Issued in the USA by the Requirements and Technical Concepts for Aviation (document RTCA SC167/DO-178B) and in Europe by the European Organization for Civil Aviation Electronics (EUROCAE document ED-12B).December 1992.

[21] IEC 61508: Parts 1-7, Functional safety: safety-related systems. 1998. (Part 3 is concerned with software).

[22] ISO/IEC 15408: 1999 Information technology. Security techniques. Evaluation criteria for IT security.

[23] J Barnes, High Integrity Software - the SPARK Approach to Safety and Security. Addison-Wesley. 2002.

[25] Steve Christy, *Vulnerability Type Distributions in CVE*, V1.0, 2006/10/04

[26] *ARIANE 5: Flight 501 Failure*, Report by the Inquiry Board, July 19, 1996 <http://esamultimedia.esa.int/docs/esa-x-1819eng.pdf>

[27] Hogaboom, Richard, *A Generic API Bit Manipulation in C*, Embedded Systems Programming, Vol 12, No 7, July 1999 <http://www.embedded.com/1999/9907/9907feat2.htm>

[28] Carlo Ghezzi and Mehdi Jazayeri, Programming Language Concepts, 3rd edition, ISBN-0-471-10426-4, John Wiley & Sons, 1998

[29] Lions, J. L. [ARIANE 5 Flight 501 Failure Report](http://en.wikisource.org/wiki/Ariane_501_Inquiry_Board_report). Paris, France: European Space Agency (ESA) & National Center for Space Study (CNES) Inquiry Board, July 1996.

[30] Seacord, R. *Secure Coding in C and C++*. Boston, MA: Addison-Wesley, 2005. See <http://www.cert.org/books/secure-coding> for news and errata.

[31] John David N. Dionisio. Type Checking. <http://myweb.lmu.edu/dondi/share/pl/type-checking-v02.pdf>

[32] MISRA Limited. "[MISRA C](http://www.misra.org.uk/): 2012 Guidelines for the Use of the C Language in Critical Systems." Warwickshire, UK: MIRA Limited, March 2013 (ISBN 978-1-906400-10-1 and 978-1-906400-11-8).

[33] The Common Weakness Enumeration (CWE) Initiative, MITRE Corporation, (<http://cwe.mitre.org/>)

[34] Goldberg, David, *What Every Computer Scientist Should Know About Floating-Point Arithmetic*, ACM Computing Surveys, vol 23, issue 1 (March 1991), ISSN 0360-0300, pp 5-48.

[35] IEEE Standards Committee 754. IEEE Standard for Binary Floating-Point Arithmetic, ANSI/IEEE Standard 754-2008. Institute of Electrical and Electronics Engineers, New York, 2008.

[36] Robert W. Sebesta, Concepts of Programming Languages, 8th edition, ISBN-13: 978-0-321-49362-0, ISBN-10: 0-321-49362-1, Pearson Education, Boston, MA, 2008

[37] Bo Einarsson, ed. Accuracy and Reliability in Scientific Computing, SIAM, July 2005 <http://www.nsc.liu.se/wg25/book>

[38] GAO Report, Patriot *Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia*, B-247094, Feb. 4, 1992, <http://archive.gao.gov/t2pbat6/145960.pdf>

[39] Robert Skeel, *Roundoff Error Cripples Patriot Missile*, SIAM News, Volume 25, Number 4, July 1992, page 11, <http://www.siam.org/siamnews/general/patriot.htm>

[40] CERT. *CERT C++ Secure Coding Standard*.  [https://www.securecoding.cert.org/confluence/pages/viewpage.action?pageId=637](https://www.securecoding.cert.org/confluence/pages/viewpage.action?pageId=637%20) (2009).

[41] Holzmann, Garard J., Computer, vol. 39, no. 6, pp 95-97, Jun., 2006, *The Power of 10: Rules for Developing Safety-Critical Code*

[42] P. V. Bhansali, A systematic approach to identifying a safe subset for safety-critical software, ACM SIGSOFT Software Engineering Notes, v.28 n.4, July 2003

[43] Ada 95 Quality and Style Guide, SPC-91061-CMC, version 02.01.01. Herndon, Virginia: Software Productivity Consortium, 1992. Available from: <http://www.adaic.org/docs/95style/95style.pdf>

[44] Ghassan, A., & Alkadi, I. (2003). Application of a Revised DIT Metric to Redesign an OO Design. *Journal of Object Technology* , 127-134.

[45] Subramanian, S., Tsai, W.-T., & Rayadurgam, S. (1998). Design Constraint Violation Detection in Safety-Critical Systems. The 3rd IEEE International Symposium on High-Assurance Systems Engineering , 109 - 116.

[46] Lundqvist, K and Asplund, L., “*A Formal Model of a Run-Time Kernel for Ravenscar*”, The 6th International Conference on Real-Time Computing Systems and Applications – RTCSA 1999

[47] ISO/IEC TS 17961, *Information technology – Programming languages, their environments and system software interfaces – C secure coding rules*

[48] GNU Project. GCC Bugs “Non-bugs” http://gcc.gnu.org/bugs.html#nonbugs\_c (2009).

# Index

LHS (left-hand side), 22

1. Integer types, Floating types and Pointer types are collectively called scalar types in the C Standard [↑](#footnote-ref-1)
2. An alias is a variable or formal parameter that refers to the same location as another variable or formal parameter. [↑](#footnote-ref-2)
3. Ada has developed a standard for interfacing with C. Fortran has included a Clause 15 that explains how to call C functions. [↑](#footnote-ref-3)
4. This has been addressed by WG 14 in an optionally normative annex in the current working paper [↑](#footnote-ref-4)
5. The first edition should not be used or quoted in this work. [↑](#footnote-ref-5)