Effective STL

50 Specific Ways to Improve Your Use of the Standard Template Library

Scott Meyers

ADDISON–WESLEY PROFESSIONAL COMPUTING SERIES

Effective STL

Addison-Wesley Professional Computing Series

Brian W. Kernighan, Consulting Editor

Matthew H. Austern, *Generic Programming and the STL: Using and Extending the C++ Standard Template Library*
David R. Butenhof, *Programming with POSIX® Threads*
Brent Callaghan, *NFS Illustrated*
Tom Cargill, *C++ Programming Style*
William R. Cheswick/Steven M. Bellovin/Aviel D. Rubin, *Firewalls and Internet Security, Second Edition: Repelling the Wily Hacker*
David A. Curry, *UNIX® System Security: A Guide for Users and System Administrators*
Stephen C. Dewhurst, *C++ Gotchas: Avoiding Common Problems in Coding and Design*
Dan Farmer/Wietse Venema, *Forensic Discovery*
Erich Gamma/Richard Helm/Ralph Johnson/John Vlissides, *Design Patterns: Elements of Reusable Object-Oriented Software*
Erich Gamma/Richard Helm/Ralph Johnson/John Vlissides, *Design Patterns CD: Elements of Reusable Object-Oriented Software*
Peter Haggar, *Practical Java™ Programming Language Guide*
David R. Hanson, *C Interfaces and Implementations: Techniques for Creating Reusable Software*
Mark Harrison/Michael McLennan, *Effective Tcl/Tk Programming: Writing Better Programs with Tcl and Tk*
Michi Henning/Steve Vinoski, *Advanced CORBA® Programming with C++*
Brian W. Kernighan/Rob Pike, *The Practice of Programming*
S. Keshav, *An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network*
John Lakos, *Large-Scale C++ Software Design*
Scott Meyers, *Effective C++ CD: 85 Specific Ways to Improve Your Programs and Designs*
Scott Meyers, *Effective C++, Third Edition: 55 Specific Ways to Improve Your Programs and Designs*
Scott Meyers, *More Effective C++: 35 New Ways to Improve Your Programs and Designs*
Scott Meyers, *Effective STL: 50 Specific Ways to Improve Your Use of the Standard Template Library*
Robert B. Murray, *C++ Strategies and Tactics*
David R. Musser/Gillmer J. Derge/Atul Saini, *STL Tutorial and Reference Guide, Second Edition: C++ Programming with the Standard Template Library*
John K. Ousterhout, *Tcl and the Tk Toolkit*
Craig Partridge, *Gigabit Networking*
Radia Perlman, *Interconnections, Second Edition: Bridges, Routers, Switches, and Internetworking Protocols*
Stephen A. Rago, *UNIX® System V Network Programming*
Eric S. Raymond, *The Art of UNIX Programming*
Marc J. Rochkind, *Advanced UNIX Programming, Second Edition*
Curt Schimmel, *UNIX® Systems for Modern Architectures: Symmetric Multiprocessing and Caching for Kernel Programmers*
W. Richard Stevens, *TCP/IP Illustrated, Volume 1: The Protocols*
W. Richard Stevens, *TCP/IP Illustrated, Volume 3: TCP for Transactions, HTTP, NNTP, and the UNIX® Domain Protocols*
W. Richard Stevens/Bill Fenner/Andrew M. Rudoff, *UNIX Network Programming Volume 1, Third Edition: The Sockets Networking API*
W. Richard Stevens/Stephen A. Rago, *Advanced Programming in the UNIX® Environment, Second Edition*
W. Richard Stevens/Gary R. Wright, *TCP/IP Illustrated Volumes 1-3 Boxed Set*
John Viega/Gary McGraw, *Building Secure Software: How to Avoid Security Problems the Right Way*
Gary R. Wright/W. Richard Stevens, *TCP/IP Illustrated, Volume 2: The Implementation*
Ruixi Yuan/W. Timothy Strayer, *Virtual Private Networks: Technologies and Solutions*

Effective STL

50 Specific Ways to Improve Your Use of the Standard Template Library

Scott Meyers

ADDISON-WESLEY

Boston • San Francisco • New York • Toronto • Montreal
London • Munich • Paris • Madrid
Capetown • Sydney • Tokyo • Singapore • Mexico City

This e-book reproduces in electronic form the printed book content of *Effective STL: 50 Specific Ways to Improve Your Use of the Standard Template Library,* by Scott Meyers. Copyright © 2001 by Addison-Wesley, an imprint of Pearson Education, Inc. ISBN: 0-201-74962-9.

Pearson Education, Inc
Rights and Contracts Department
501 Boylston Street, Suite 900
Boston, MA 02116
Fax (617) 671-3447

E-book ISBN 13: 978-0-321-51580-3
E-book ISBN 10: 0-321-51580-3
Second e-book release, April 2011 (essentially identical to the 13th Paper Printing).

For Woofieland.

This page intentionally left blank

Contents

This page intentionally left blank

Preface

It came without ribbons! It came without tags!
It came without packages, boxes or bags!

— Dr. Seuss, *How the Grinch Stole Christmas!*, Random House, 1957

I first wrote about the Standard Template Library in 1995, when I concluded the final Item of *More Effective C++* with a brief STL overview. I should have known better. Shortly thereafter, I began receiving mail asking when I'd write *Effective STL*.

I resisted the idea for several years. At first, I wasn't familiar enough with the STL to offer advice on it, but as time went on and my experience with it grew, this concern gave way to other reservations. There was never any question that the library represented a breakthrough in efficient and extensible design, but when it came to *using* the STL, there were practical problems I couldn't overlook. Porting all but the simplest STL programs was a challenge, not only because library implementations varied, but also because template support in the underlying compilers ranged from good to awful. STL tutorials were hard to come by, so learning "the STL way of programming" was difficult, and once that hurdle was overcome, finding comprehensible and accurate reference documentation was a challenge. Perhaps most daunting, even the smallest STL usage error often led to a blizzard of compiler diagnostics, each thousands of characters long, most referring to classes, functions, or templates not mentioned in the offending source code, almost all incomprehensible. Though I had great admiration for the STL and for the people behind it, I felt uncomfortable recommending it to practicing programmers. I wasn't sure it was *possible* to use the STL effectively.

Then I began to notice something that took me by surprise. Despite the portability problems, despite the dismal documentation, despite the compiler diagnostics resembling transmission line noise, many of my consulting clients were using the STL anyway. Furthermore, they weren't just playing with it, they were using it in production code! That was a revelation. I knew that the STL featured an elegant design, but any library for which programmers are willing to endure portability headaches, poor documentation, and incomprehensible error messages has a lot more going for it than just good design. For an increasingly large number of professional programmers, I realized, even a bad implementation of the STL was preferable to no implementation at all.

Furthermore, I knew that the situation regarding the STL would only get better. Libraries and compilers would grow more conformant with the Standard (they have), better documentation would become available (it has — consult the bibliography beginning on page 225), and compiler diagnostics would improve (for the most part, we're still waiting, but Item 49 offers suggestions for how to cope while we wait). I therefore decided to chip in and do my part for the STL movement. This book is the result: 50 specific ways to improve your use of C++'s Standard Template Library.

My original plan was to write the book in the second half of 1999, and with that thought in mind, I put together an outline. But then I changed course. I suspended work on the book and developed an introductory training course on the STL, which I then taught several times to groups of programmers. About a year later, I returned to the book, significantly revising the outline based on my experiences with the training course. In the same way that my *Effective C++* has been successful by being grounded in the problems faced by real programmers, it's my hope that *Effective STL* similarly addresses the practical aspects of STL programming — the aspects most important to professional developers.

I am always on the lookout for ways to improve my understanding of C++. If you have suggestions for new guidelines for STL programming or if you have comments on the guidelines in this book, please let me know. In addition, it is my continuing goal to make this book as accurate as possible, so for each error in this book that is reported to me — be it technical, grammatical, typographical, or otherwise — I will, in future printings, gladly add to the acknowledgments the name of the first person to bring that error to my attention. Send your suggested guidelines, your comments, and your criticisms to estl@aristeia.com.

I maintain a list of changes to this book since its first printing, including bug-fixes, clarifications, and technical updates. The list is available at the *Effective STL Errata* web site, http://www.aristeia.com/BookErrata/estl1e-errata.html.

If you'd like to be notified when I make changes to this book, I encourage you to join my mailing list. I use the list to make announcements likely to be of interest to people who follow my work on C++. For details, consult http://www.aristeia.com/MailingList/.

SCOTT DOUGLAS MEYERS
http://www.aristeia.com/

STAFFORD, OREGON
APRIL 2001

This page intentionally left blank

Acknowledgments

I had an enormous amount of help during the roughly two years it took me to make some sense of the STL, create a training course on it, and write this book. Of all my sources of assistance, two were particularly important. The first is Mark Rodgers. Mark generously volunteered to review my training materials as I created them, and I learned more about the STL from him than from anybody else. He also acted as a technical reviewer for this book, again providing observations and insights that improved virtually every Item.

The other outstanding source of information was several C++-related Usenet newsgroups, especially comp.lang.c++.moderated ("clcm"), comp.std.c++, and microsoft.public.vc.stl. For well over a decade, I've depended on the participants in newsgroups like these to answer my questions and challenge my thinking, and it's difficult to imagine what I'd do without them. I am deeply grateful to the Usenet community for their help with both this book and my prior publications on C++.

My understanding of the STL was shaped by a variety of publications, the most important of which are listed in the Bibliography. I leaned especially heavily on Josuttis' *The C++ Standard Library* [3].

This book is fundamentally a summary of insights and observations made by others, though a few of the ideas are my own. I've tried to keep track of where I learned what, but the task is hopeless, because a typical Item contains information garnered from many sources over a long period of time. What follows is incomplete, but it's the best I can do. Please note that my goal here is to summarize where *I* first learned of an idea or technique, not where the idea or technique was originally developed or who came up with it.

In Item 1, my observation that node-based containers offer better support for transactional semantics is based on section 5.11.2 of Josuttis' *The C++ Standard Library* [3]. Item 2 includes an example from Mark Rodgers on how typedefs help when allocator types are changed.

Item 5 was motivated by Reeves' *C++ Report* column, "STL Gotchas" [17]. Item 8 sprang from Item 37 in Sutter's *Exceptional C++* [8], and Kevlin Henney provided important details on how containers of auto_ptrs fail in practice. In Usenet postings, Matt Austern provided examples of when allocators are useful, and I include his examples in Item 11. Item 12 is based on the discussion of thread safety at the SGI STL web site [21]. The material in Item 13 on the performance implications of reference counting in a multithreaded environment is drawn from Sutter's writings on this topic [20]. The idea for Item 15 came from Reeves' *C++ Report* column, "Using Standard string in the Real World, Part 2," [18]. In Item 16, Mark Rodgers came up with the technique I show for having a C API write data directly into a vector. Item 17 includes information from Usenet postings by Siemel Naran and Carl Barron. I stole Item 18 from Sutter's *C++ Report* column, "When Is a Container Not a Container?" [12]. In Item 20, Mark Rodgers contributed the idea of transforming a pointer into an object via a dereferencing functor, and Scott Lewandowski came up with the version of DereferenceLess I present. Item 21 originated in a Doug Harrison posting to microsoft.public.vc.stl, but the decision to restrict the focus of that Item to equality was mine. I based Item 22 on Sutter's *C++ Report* column, "Standard Library News: sets and maps" [13]; Matt Austern helped me understand the status of the Standardization Committee's Library Issue #103. Item 23 was inspired by Austern's *C++ Report* article, "Why You Shouldn't Use set — and What to Use Instead" [15]; David Smallberg provided a neat refinement for my implementation of DataCompare. My description of Dinkumware's hashed containers is based on Plauger's *C/C++ Users Journal* column, "Hash Tables" [16]. Mark Rodgers doesn't agree with the overall advice of Item 26, but an early motivation for that Item was his observation that some container member functions accept only arguments of type iterator. My treatment of Item 29 was motivated and informed by Usenet discussions involving Matt Austern and James Kanze; I was also influenced by Kreft and Langer's *C++ Report* article, "A Sophisticated Implementation of User-Defined Inserters and Extractors" [25]. Item 30 is due to a discussion in section 5.4.2 of Josuttis' *The C++ Standard Library* [3]. In Item 31, Marco Dalla Gasperina contributed the example use of nth_element to calculate medians, and use of that algorithm for finding percentiles comes straight out of section 18.7.1 of Stroustrup's *The C++ Programming Language* [7]. Item 32 was influenced by the material in section 5.6.1 of Josuttis' *The C++ Standard Library* [3]. Item 35 originated in Austern's *C++ Report* column "How to Do Case-Insensitive String Comparison" [11], and James Kanze's and John Potter's clcm postings helped me refine my understanding of the issues involved. Stroustrup's implementation for copy_if, which I show in Item 36, is from section 18.6.1 of his *The C++ Programming Language* [7]. Item 39 was largely motivated by the publications of Josuttis, who has written about "stateful predicates" in his *The C++ Standard Library* [3], in Standard Library Issue #92, and in his *C++ Report* article, "Predicates vs. Function Objects" [14]. In my treatment, I use his example and recommend a solution he has proposed, though the use of the term "pure function" is my own. Matt Austern confirmed my suspicion in Item 41 about the history of the terms mem_fun and mem_fun_ref. Item 42 can be traced to a lecture I got from Mark Rodgers when I considered violating that guideline. Mark Rodgers is also responsible for the insight in Item 44 that non-member searches over maps and multimaps examine both components of each pair, while member searches examine only the first (key) component. Item 45 contains information from various clcm contributors, including John Potter, Marcin Kasperski, Pete Becker, Dennis Yelle, and David Abrahams. David Smallberg alerted me to the utility of equal_range in performing equivalence-based searches and counts over sorted sequence containers. Andrei Alexandrescu helped me understand the conditions under which "the reference-to-reference problem" I describe in Item 50 arises, and I modeled my example of the problem on a similar example provided by Mark Rodgers at the Boost Web Site [22].

Credit for the material in Appendix A goes to Matt Austern, of course. I'm grateful that he not only gave me permission to include it in this book, he also tweaked it to make it even better than the original.

Good technical books require a thorough pre-publication vetting, and I was fortunate to benefit from the insights of an unusually talented group of technical reviewers. Brian Kernighan and Cliff Green offered early comments on a partial draft, and complete versions of the manuscript were scrutinized by Doug Harrison, Brian Kernighan, Tim Johnson, Francis Glassborow, Andrei Alexandrescu, David Smallberg, Aaron Campbell, Jared Manning, Herb Sutter, Stephen Dewhurst, Matt Austern, Gillmer Derge, Aaron Moore, Thomas Becker, Victor Von, and, of course, Mark Rodgers. Katrina Avery did the copyediting.

One of the most challenging parts of preparing a book is finding good technical reviewers. I thank John Potter for introducing me to Jared Manning and Aaron Campbell.

Herb Sutter kindly agreed to act as my surrogate in compiling, running, and reporting on the behavior of some STL test programs under a beta version of Microsoft's Visual Studio .NET, while Leor Zolman undertook the herculean task of testing all the code in this book. Any errors that remain are my fault, of course, not Herb's or Leor's.

Angelika Langer opened my eyes to the indeterminate status of some aspects of STL function objects. This book has less to say about function objects than it otherwise might, but what it does say is more likely to remain true. At least I hope it is.

This printing of the book is better than earlier printings, because I was able to address problems identified by the following sharp-eyed readers: Jon Webb, Michael Hawkins, Derek Price, Jim Scheller, Carl Manaster, Herb Sutter, Albert Franklin, George King, Dave Miller, Harold Howe, John Fuller, Tim McCarthy, John Hershberger, Igor Mikolic-Torreira, Stephan Bergmann, Robert Allan Schwartz, John Potter, David Grigsby, Sanjay Pattni, Jesper Andersen, Jing Tao Wang, André Blavier, Dan Schmidt, Bradley White, Adam Petersen, Wayne Goertel, Gabriel Netterdag, Jason Kenny, Scott Blachowicz, Seyed H. Haeri, Gareth McCaughan, Giulio Agostini, Fraser Ross, Wolfram Burkhardt, Keith Stanley, Leor Zolman, Chan Ki Lok, Motti Abramsky, Kevlin Henney, Stefan Kuhlins, Phillip Ngan, Jim Phillips, Ruediger Dreier, Guru Chandar, Charles Brockman, Day Barr, Eric Niebler, Sharad Kala, Declan Moran, Nick de Smith, David Callaway, Shlomi Frank, Andrea Griffini, Hans Eckardt, David Smallberg, Matt Page, Andy Fyfe, Vincent Stojanov, Randy Parker, Thomas Schell, Cameron Mac Minn, Mark Davis, Giora Unger, Julie Nahil, Martin Rottinger, Neil Henderson, Andrew Savige, and Molly Sharp. I'm grateful for their help in improving *Effective STL*.

My collaborators at Addison-Wesley included John Wait (my editor and now a senior VP), Alicia Carey and Susannah Buzard (his assistants n and n+1), John Fuller (the production coordinator), Karin Hansen (the cover designer), Jason Jones (all-around technical guru, especially with respect to the demonic software spewed forth by Adobe), Marty Rabinowitz (their boss, but he works, too), and Curt Johnson, Chanda Leary-Coutu, and Robin Bruce (all marketing people, but still very nice).

Abbi Staley made Sunday lunches a routinely pleasurable experience.

As she has for the six books and one CD that came before it, my wife, Nancy, tolerated the demands of my research and writing with her usual forbearance and offered me encouragement and support when I needed it most. She never fails to remind me that there's more to life than C++ and software.

And then there's our dog, Persephone. As I write this, it is her sixth birthday. Tonight, she and Nancy and I will visit Baskin-Robbins for ice cream. Persephone will have vanilla. One scoop. In a cup. To go.

Introduction

You're already familiar with the STL. You know how to create containers, iterate over their contents, add and remove elements, and apply common algorithms, such as find and sort. But you're not satisfied. You can't shake the sensation that the STL offers more than you're taking advantage of. Tasks that should be simple aren't. Operations that should be straightforward leak resources or behave erratically. Procedures that should be efficient demand more time or memory than you're willing to give them. Yes, you know how to use the STL, but you're not sure you're using it *effectively*.

I wrote this book for you.

In *Effective STL*, I explain how to combine STL components to take full advantage of the library's design. Such information allows you to develop simple, straightforward solutions to simple, straightforward problems, and it also helps you design elegant approaches to more complicated problems. I describe common STL usage errors, and I show you how to avoid them. That helps you dodge resource leaks, code that won't port, and behavior that is undefined. I discuss ways to optimize your code, so you can make the STL perform like the fast, sleek machine it is intended to be.

The information in this book will make you a better STL programmer. It will make you a more productive programmer. And it will make you a happier programmer. Using the STL is fun, but using it effectively is outrageous fun, the kind of fun where they have to drag you away from the keyboard, because you just can't believe the good time you're having. Even a cursory glance at the STL reveals that it is a wondrously cool library, but the coolness runs broader and deeper than you probably imagine. One of my primary goals in this book is to convey to you just how amazing the library is, because in the nearly 30 years I've been programming, I've never seen anything like the STL. You probably haven't either.

Defining, Using, and Extending the STL

There is no official definition of "the STL," and different people mean different things when they use the term. In this book, "the STL" means the parts of C++'s Standard Library that work with iterators. That includes the standard containers (including string), parts of the iostream library, function objects, and algorithms. It excludes the standard container adapters (stack, queue, and priority_queue) as well as the containers bitset and valarray, because they lack iterator support. It doesn't include arrays, either. True, arrays support iterators in the form of pointers, but arrays are part of the C++ *language*, not the library.

Technically, my definition of the STL excludes extensions of the standard C++ library, notably hashed containers, singly linked lists, ropes, and a variety of nonstandard function objects. Even so, an effective STL programmer needs to be aware of such extensions, so I mention them where it's appropriate. Indeed, Item 25 is devoted to an overview of nonstandard hashed containers. They're not in the STL now, but something similar to them is almost certain to make it into the next version of the standard C++ library, and there's value in glimpsing the future.

One of the reasons for the existence of STL extensions is that the STL is a library designed to be extended. In this book, however, I focus on *using* the STL, not on adding new components to it. You'll find, for example, that I have little to say about writing your own algorithms, and I offer no guidance at all on writing new containers and iterators. I believe that it's important to master what the STL already provides before you embark on increasing its capabilities, so that's what I focus on in *Effective STL*. When you decide to create your own STLesque components, you'll find advice on how to do it in books like Josuttis' *The C++ Standard Library* [3] and Austern's *Generic Programming and the STL* [4]. One aspect of STL extension I *do* discuss in this book is writing your own function objects. You can't use the STL effectively without knowing how to do that, so I've devoted an entire chapter to the topic (Chapter 6).

Citations

The references to the books by Josuttis and Austern in the preceding paragraph demonstrate how I handle bibliographic citations. In general, I try to mention enough of a cited work to identify it for people who are already familiar with it. If you already know about these authors' books, for example, you don't have to turn to the Bibliography to find out that [3] and [4] refer to books you already know. If you're not familiar with a publication, of course, the Bibliography (which begins on page 225) gives you a full citation.

I cite three works often enough that I generally leave off the citation number. The first of these is the International Standard for C++ [5], which I usually refer to as simply "the Standard." The other two are my earlier books on C++, *Effective C++* [1] and *More Effective C++* [2].

The STL and Standards

I refer to the C++ Standard frequently, because *Effective STL* focuses on portable, standard-conformant C++. In theory, everything I show in this book will work with every C++ implementation. In practice, that isn't true. Shortcomings in compiler and STL implementations conspire to prevent some valid code from compiling or from behaving the way it's supposed to. Where that is commonly the case, I describe the problems, and I explain how you can work around them.

Sometimes, the easiest workaround is to use a different STL implementation. Appendix B gives an example of when this is the case. In fact, the more you work with the STL, the more important it becomes to distinguish between your *compilers* and your *library implementations*. When programmers run into difficulties trying to get legitimate code to compile, it's customary for them to blame their compilers, but with the STL, compilers can be fine, while STL implementations are faulty. To emphasize the fact that you are dependent on both your compilers and your library implementations, I refer to your *STL platforms*. An STL platform is the combination of a particular compiler and a particular STL implementation. In this book, if I mention a compiler problem, you can be sure that I mean it's the compiler that's the culprit. However, if I refer to a problem with your STL platform, you should interpret that as "maybe a compiler bug, maybe a library bug, possibly both."

I generally refer to your "compilers" — *plural*. That's an outgrowth of my longstanding belief that you improve the quality (especially the portability) of your code if you ensure that it works with more than one compiler. Furthermore, using multiple compilers generally makes it easier to unravel the Gordian nature of error messages arising from improper use of the STL. (Item 49 is devoted to approaches to decoding such messages.)

Another aspect of my emphasis on standard-conforming code is my concern that you avoid constructs with undefined behavior. Such constructs may do anything at runtime. Unfortunately, this means they may do precisely what you want them to, and that can lead to a false sense of security. Too many programmers assume that undefined behavior always leads to an obvious problem, e.g., a segmentation fault or other catastrophic failure. The results of undefined behavior can actually be much more subtle, e.g., corruption of rarely-referenced data. They can also vary across program runs. I find that a good working definition of undefined behavior is "works for me, works for you, works during development and QA, but blows up in your most important customer's face." It's important to avoid undefined behavior, so I point out common situations where it can arise. You should train yourself to be alert for such situations.

Reference Counting

It's close to impossible to discuss the STL without mentioning reference counting. As you'll see in Items 7 and 33, designs based on containers of pointers almost invariably lead to reference counting. In addition, many string implementations are internally reference counted, and, as Item 15 explains, this may be an implementation detail you can't afford to ignore. In this book, I assume that you are familiar with the basics of reference counting. If you're not, most intermediate and advanced C++ texts cover the topic. In *More Effective C++*, for example, the relevant material is in Items 28 and 29. If you don't know what reference counting is and you have no inclination to learn, don't worry. You'll get through this book just fine, though there may be a few sentences here and there that won't make as much sense as they otherwise would.

string and wstring

Whatever I say about string applies equally well to its wide-character counterpart, wstring. Similarly, any time I refer to the relationship between string and char or char*, the same is true of the relationship between wstring and wchar_t or wchar_t*. In other words, just because I don't explicitly mention wide-character strings in this book, don't assume that the STL fails to support them. It supports them as well as char-based strings. It has to. Both string and wstring are instantiations of the same template, basic_string.

Terms, Terms, Terms

This is not an introductory book on the STL, so I assume you know the fundamentals. Still, the following terms are sufficiently important that I feel compelled to review them:

- vector, string, deque, and list are known as the *standard sequence containers*. The *standard associative containers* are set, multiset, map, and multimap.
- Iterators are divided into five categories, based on the operations they support. Very briefly, *input iterators* are read-only iterators where each iterated location may be read only once. *Output iterators* are write-only iterators where each iterated location may be written only once. Input and output iterators are modeled on reading and writing input and output streams (e.g., files). It's thus unsurprising that the most common manifestations of input and output iterators are istream_iterators and ostream_iterators, respectively.

  Forward iterators have the capabilities of both input and output iterators, but they can read or write a single location repeatedly. They don't support operator--, so they can move only forward with any degree of efficiency. All standard STL containers support iterators that are more powerful than forward iterators, but, as you'll see in Item 25, one design for hashed containers yields forward iterators. Containers for singly linked lists (considered in Item 50) also offer forward iterators.

  Bidirectional iterators are just like forward iterators, except they can go backward as easily as they go forward. The standard associative containers all offer bidirectional iterators. So does list.

  Random access iterators do everything bidirectional iterators do, but they also offer "iterator arithmetic," i.e., the ability to jump forward or backward in a single step. vector, string, and deque each provide random access iterators. Pointers into arrays act as random access iterators for the arrays.
- Any class that overloads the function call operator (i.e., operator()) is a *functor class*. Objects created from such classes are known as *function objects* or *functors*. Most places in the STL that work with function objects work equally well with real functions, so I often use the term "function objects" to mean both C++ functions as well as true function objects.
- The functions bind1st and bind2nd are known as *binders*.

A revolutionary aspect of the STL is its complexity guarantees. These guarantees bound the amount of work any STL operation is allowed to perform. This is wonderful, because it can help you determine the relative efficiency of different approaches to the same problem, regardless of the STL platform you're using. Unfortunately, the terminology behind the complexity guarantees can be confusing if you haven't been formally introduced to the jargon of computer science. Here's a quick primer on the complexity terms I use in this book. Each refers to how long it takes to do something as a function of *n*, the number of elements in a container or range.

- An operation that runs in *constant time* has performance that is unaffected by changes in *n*. For example, inserting an element into a list is a constant-time operation. Regardless of whether the list has one element or one million, the insertion takes about the same amount of time.

  Don't take the term "constant time" too literally. It doesn't mean that the amount of time it takes to do something is literally constant, it just means that it's unaffected by *n*. For example, two STL platforms might take dramatically different amounts of time to perform the same "constant-time" operation. This could happen if one library has a much more sophisticated implementation than another or if one compiler performs substantially more aggressive optimization.

  A variant of constant time complexity is *amortized constant time*. Operations that run in amortized constant time are usually constant-time operations, but occasionally they take time that depends on *n*. Amortized constant time operations *typically* run in constant time.

- An operation that runs in *logarithmic time* needs more time to run as *n* gets larger, but the time it requires grows at a rate proportional to the logarithm of *n*. For example, an operation on a million items would be expected to take only about three times as long as on a hundred items, because $\log n^3 = 3 \log n$. Most search operations on associative containers (e.g., set::find) are logarithmic-time operations.

- The time needed to perform an operation that runs in *linear time* increases at a rate proportional to increases in *n*. The standard algorithm count runs in linear time, because it has to look at every element of the range it's given. If the range triples in size, it has to do three times as much work, and we'd expect it to take about three times as long to do it.

As a general rule, a constant-time operation runs faster than one requiring logarithmic time, and a logarithmic-time operation runs faster than one whose performance is linear. This is always true when *n* gets big enough, but for relatively small values of *n*, it's sometimes possible for an operation with a worse theoretical complexity to outperform an operation with a better theoretical complexity. If you'd like to know more about STL complexity guarantees, turn to Josuttis' *The C++ Standard Library* [3].

As a final note on terminology, recall that each element in a map or multimap has two components. I generally call the first component the *key* and the second component the *value*. Given

```
map<string, double> m;
```

for example, the string is the key and the double is the value.

Code Examples

This book is filled with example code, and I explain each example when I introduce it. Still, it's worth knowing a few things in advance.

You can see from the map example above that I routinely omit #includes and ignore the fact that STL components are in namespace std. When defining the map m, I could have written this,

```
#include <map>
#include <string>

using std::map;
using std::string;

map<string, double> m;
```

but I prefer to save us both the noise.

When I declare a formal type parameter for a template, I use typename instead of class. That is, instead of writing this,

```
template<class T>
class Widget { ... };
```

I write this:

```
template<typename T>
class Widget { ... };
```

In this context, class and typename mean exactly the same thing, but I find that typename more clearly expresses what I usually want to say: that *any* type will do; T need not be a class. If you prefer to use class to declare type parameters, go right ahead. Whether to use typename or class in this context is purely a matter of style.

It is not a matter of style in a different context. To avoid potential parsing ambiguities (the details of which I'll spare you), you are required to use typename to precede type names that are dependent on formal type parameters. Such types are known as *dependent types*, and an example will help clarify what I'm talking about. Suppose you'd like to write a template for a function that, given an STL container, returns whether the last element in the container is greater than the first element. Here's one way to do it:

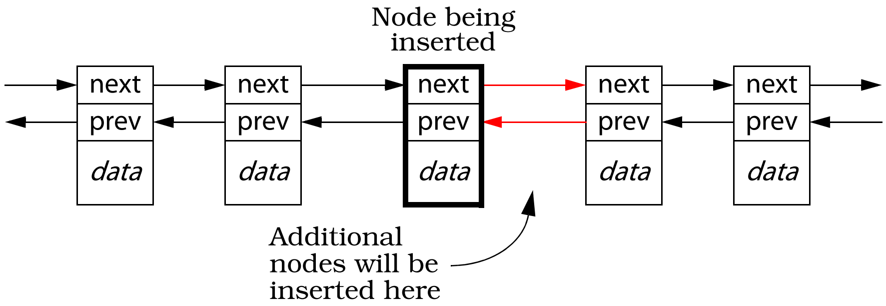
```
template<typename C>
bool lastGreaterThanFirst(const C& container)
{
   if (container.empty()) return false;

   typename C::const_iterator begin(container.begin());
   typename C::const_iterator end(container.end());

   return *--end > *begin;
}
```

In this example, the local variables begin and end are of type C::const_iterator. const_iterator is a type that is dependent on the formal type parameter C. Because C::const_iterator is a dependent type, you are required to precede it with the word typename. (Some compilers incorrectly accept the code without the typenames, but such code isn't portable.)

I hope you've noticed my use of color in the examples above. It's there to focus your attention on parts of the code that are particularly important. Often, I highlight the differences between related examples, such as when I showed the two possible ways to declare the parameter T in the Widget example. This use of color to call out especially noteworthy parts of examples carries over to diagrams, too. For instance, this diagram from Item 5 uses color to identify the two pointers that are affected when a new element is inserted into a list:



I also use color for chapter numbers, but such use is purely gratuitous. This being my first two-color book, I hope you'll forgive me a little chromatic exuberance.

Two of my favorite parameter names are lhs and rhs. They stand for "left-hand side" and "right-hand side," respectively, and I find them especially useful when declaring operators. Here's an example from Item 19:

```
class Widget { ... };

bool operator==(const Widget& lhs, const Widget& rhs);
```

When this function is called in a context like this,

```
if (x == y) ...                          // assume x and y are Widgets
```

x, which is on the left-hand side of the "==", is known as lhs inside operator==, and y is known as rhs.

As for the class name Widget, that has nothing to do with GUIs or toolkits. It's just the name I use for "some class that does something." Sometimes, as on page 7, Widget is a class template instead of a class. In such cases, you may find that I still refer to Widget as a class, even though it's really a template. Such sloppiness about the difference between classes and class templates, structs and struct templates, and functions and function templates hurts no one as long as there is no ambiguity about what is being discussed. In cases where it could be confusing, I do distinguish between templates and the classes, structs, and functions they generate.

Efficiency Items

I considered including a chapter on efficiency in *Effective STL*, but I ultimately decided that the current organization was preferable. Still, a number of Items focus on minimizing space and runtime demands. For your performance-enhancing convenience, here is the table of contents for the virtual chapter on efficiency:

The Guidelines in *Effective STL*

The guidelines that make up the 50 Items in this book are based on the insights and advice of the world's most experienced STL programmers. These guidelines summarize things you should almost always do — or almost always avoid doing — to get the most out of the Standard Template Library. At the same time, they're just guidelines. Under some conditions, it makes sense to violate them. For example, the title of Item 7 tells you to invoke `delete` on `new`ed pointers in a container before the container is destroyed, but the text of that Item makes clear that this applies only when the objects pointed to by those pointers should go away when the container does. This is often the case, but it's not universally true. Similarly, the title of Item 35 beseeches you to use STL algorithms to perform simple case-insensitive string comparisons, but the text of the Item points out that in some cases, you'll be better off using a function that's not only outside the STL, it's not even part of standard C++!

Only you know enough about the software you're writing, the environment in which it will run, and the context in which it's being created to determine whether it's reasonable to violate the guidelines I present. Most of the time, it won't be, and the discussions that accompany each Item explain why. In a few cases, it will. Slavish devotion to the guidelines isn't appropriate, but neither is cavalier disregard. Before venturing off on your own, you should make sure you have a good reason.

1 Containers

Sure, the STL has iterators, algorithms, and function objects, but for most C++ programmers, it's the containers that stand out. More powerful and flexible than arrays, they grow (and often shrink) dynamically, manage their own memory, keep track of how many objects they hold, bound the algorithmic complexity of the operations they support, and much, much more. Their popularity is easy to understand. They're simply better than their competition, regardless of whether that competition comes from containers in other libraries or is a container type you'd write yourself. STL containers aren't just good. They're *really* good.

This chapter is devoted to guidelines applicable to all the STL containers. Later chapters focus on specific container types. The topics addressed here include selecting the appropriate container given the constraints you face; avoiding the delusion that code written for one container type is likely to work with other container types; the significance of copying operations for objects in containers; difficulties that arise when pointers or auto_ptrs are stored in containers; the ins and outs of erasing; what you can and cannot accomplish with custom allocators; tips on how to maximize efficiency; and considerations for using containers in a threaded environment.

That's a lot of ground to cover, but don't worry. The Items break it down into bite-sized chunks, and along the way, you're almost sure to pick up several ideas you can apply to your code *now*.

Item 1: Choose your containers with care.

You know that C++ puts a variety of containers at your disposal, but do you realize just how varied that variety is? To make sure you haven't overlooked any of your options, here's a quick review.

- The **standard STL sequence containers**, vector, string, deque, and list.

- The **standard STL associative containers**, set, multiset, map, and multimap.
- The **nonstandard sequence containers** slist and rope. slist is a singly linked list, and rope is essentially a heavy-duty string. (A "rope" is a heavy-duty "string." Get it?) You'll find a brief overview of these nonstandard (but commonly available) containers in Item 50.
- The **nonstandard associative containers** hash_set, hash_multiset, hash_map, and hash_multimap. I examine these widely available hash-table-based variants on the standard associative containers in Item 25.
- **vector<char> as a replacement for string**. Item 13 describes the conditions under which such a replacement might make sense.
- **vector as a replacement for the standard associative containers**. As Item 23 makes clear, there are times when vector can outperform the standard associative containers in both time and space.
- Several **standard non-STL containers**, including arrays, bitset, valarray, stack, queue, and priority_queue. Because these are non-STL containers, I have little to say about them in this book, though Item 16 mentions a case where arrays are preferable to STL containers and Item 18 explains why bitset may be better than vector<bool>. It's also worth bearing in mind that arrays can be used with STL algorithms, because pointers can be used as array iterators.

That's a panoply of options, and it's matched in richness by the range of considerations that should go into choosing among them. Unfortunately, most discussions of the STL take a fairly narrow view of the world of containers, ignoring many issues relevant to selecting the one that is most appropriate. Even the Standard gets into this act, offering the following guidance for choosing among vector, deque, and list:

> vector, list, and deque offer the programmer different complexity trade-offs and should be used accordingly. vector is the type of sequence that should be used by default. list should be used when there are frequent insertions and deletions from the middle of the sequence. deque is the data structure of choice when most insertions and deletions take place at the beginning or at the end of the sequence.

If your primary concern is algorithmic complexity, I suppose this constitutes reasonable advice, but there is so much more to be concerned with.

In a moment, we'll examine some of the important container-related issues that complement algorithmic complexity, but first I need to introduce a way of categorizing the STL containers that isn't discussed as often as it should be. That is the distinction between contiguous-memory containers and node-based containers.

Contiguous-memory containers (also known as *array-based containers*) store their elements in one or more (dynamically allocated) chunks of memory, each chunk holding more than one container element. If a new element is inserted or an existing element is erased, other elements in the same memory chunk have to be shifted up or down to make room for the new element or to fill the space formerly occupied by the erased element. This kind of movement affects both performance (see Items 5 and 14) and exception safety (as we'll soon see). The standard contiguous-memory containers are vector, string, and deque. The nonstandard rope is also a contiguous-memory container.

Node-based containers store only a single element per chunk of (dynamically allocated) memory. Insertion or erasure of a container element affects only pointers to nodes, not the contents of the nodes themselves, so element values need not be moved when something is inserted or erased. Containers representing linked lists, such as list and slist, are node-based, as are all the standard associative containers. (They're typically implemented as balanced trees.) The nonstandard hashed containers use varying node-based implementations, as you'll see in Item 25.

With this terminology out of the way, we're ready to sketch some of the questions most relevant when choosing among containers. In this discussion, I omit consideration of non-STL-like containers (e.g., arrays, bitsets, etc.), because this is, after all, a book on the STL.

- *Do you need to be able to insert a new element at an arbitrary position in the container?* If so, you need a sequence container; associative containers won't do.
- *Do you care how elements are ordered in the container?* If not, a hashed container becomes a viable choice. Otherwise, you'll want to avoid hashed containers.
- *Must the container be part of standard C++?* If so, that eliminates hashed containers, slist, and rope.
- *What category of iterators do you require?* If they must be random access iterators, you're technically limited to vector, deque, and string, but you'd probably want to consider rope, too. (See Item 50

for information on rope.) If bidirectional iterators are required, you must avoid slist (see Item 50) as well as one common implementation of the hashed containers (see Item 25).

- *Is it important to avoid movement of existing container elements when insertions or erasures take place?* If so, you'll need to stay away from contiguous-memory containers (see Item 5).
- *Does the data in the container need to be layout-compatible with C?* If so, you're limited to vectors (see Item 16).
- *Is lookup speed a critical consideration?* If so, you'll want to look at hashed containers (see Item 25), sorted vectors (see Item 23), and the standard associative containers — probably in that order.
- *Do you mind if the underlying container uses reference counting?* If so, you'll want to steer clear of string, because many string implementations are reference-counted (see Item 13). You'll need to avoid rope, too, because the definitive rope implementation is based on reference counting (see Item 50). You have to represent your strings somehow, of course, so you'll want to consider vector<char>.
- *Do you need transactional semantics for insertions and erasures?* That is, do you require the ability to reliably roll back insertions and erasures? If so, you'll want to use a node-based container. If you need transactional semantics for multiple-element insertions (e.g., the range form — see Item 5), you'll want to choose list, because list is the only standard container that offers transactional semantics for multiple-element insertions. Transactional semantics are particularly important for programmers interested in writing exception-safe code. (Transactional semantics can be achieved with contiguous-memory containers, too, but there is a performance cost, and the code is not as straightforward. To learn more about this, consult Item 17 of Sutter's *Exceptional C++* [8].)
- *Do you need to minimize iterator, pointer, and reference invalidation?* If so, you'll want to use node-based containers, because insertions and erasures on such containers never invalidate iterators, pointers, or references (unless they point to an element you are erasing). In general, insertions or erasures on contiguous-memory containers may invalidate all iterators, pointers, and references into the container.
- *Do you care if using swap on containers invalidates iterators, pointers, or references?* If so, you'll need to avoid string, because string is alone in the STL in invalidating iterators, pointers, and references during swaps.

- *Would it be helpful to have a sequence container with random access iterators where pointers and references to the data are not invalidated as long as nothing is erased and insertions take place only at the ends of the container?* This is a very special case, but if it's *your* case, deque is the container of your dreams. (Interestingly, deque's iterators *may* be invalidated when insertions are made only at the ends of the container. deque is the only standard STL container whose iterators may be invalidated without also invalidating its pointers and references.)

These questions are hardly the end of the matter. For example, they don't take into account the varying memory allocation strategies employed by the different container types. (Items 10 and 14 discuss some aspects of such strategies.) Still, they should be enough to convince you that, unless you have no interest in element ordering, standards conformance, iterator capabilities, layout compatibility with C, lookup speed, behavioral anomalies due to reference counting, the ease of implementing transactional semantics, or the conditions under which iterators are invalidated, you have more to think about than simply the algorithmic complexity of container operations. Such complexity is important, of course, but it's far from the entire story.

The STL gives you lots of options when it comes to containers. If you look beyond the bounds of the STL, there are even more options. Before choosing a container, be sure to consider *all* your options. A "default container"? I don't think so.

Item 2: Beware the illusion of container-independent code.

The STL is based on generalization. Arrays are generalized into containers and parameterized on the types of objects they contain. Functions are generalized into algorithms and parameterized on the types of iterators they use. Pointers are generalized into iterators and parameterized on the type of objects they point to.

That's just the beginning. Individual container types are generalized into sequence and associative containers, and similar containers are given similar functionality. Standard contiguous-memory containers (see Item 1) offer random-access iterators, while standard node-based containers (again, see Item 1) provide bidirectional iterators. Sequence containers support push_front and/or push_back, while associative containers don't. Associative containers offer logarithmic-time lower_bound, upper_bound, and equal_range member functions, but sequence containers don't.

With all this generalization going on, it's natural to want to join the movement. This sentiment is laudable, and when you write your own containers, iterators, and algorithms, you'll certainly want to pursue it. Alas, many programmers try to pursue it in a different manner. Instead of committing to particular types of containers in their software, they try to generalize the notion of a container so that they can use, say, a vector, but still preserve the option of replacing it with something like a deque or a list later — all without changing the code that uses it. That is, they strive to write *container-independent code*. This kind of generalization, well-intentioned though it is, is almost always misguided.

Even the most ardent advocate of container-independent code soon realizes that it makes little sense to try to write software that will work with both sequence and associative containers. Many member functions exist for only one category of container, e.g., only sequence containers support push_front or push_back, and only associative containers support count and lower_bound, etc. Even such basics as insert and erase have signatures and semantics that vary from category to category. For example, when you insert an object into a sequence container, it stays where you put it, but if you insert an object into an associative container, the container moves the object to where it belongs in the container's sort order. For another example, the form of erase taking an iterator returns a new iterator when invoked on a sequence container, but it returns nothing when invoked on an associative container. (Item 9 gives an example of how this can affect the code you write.)

Suppose, then, you aspire to write code that can be used with the most common sequence containers: vector, deque, and list. Clearly, you must program to the intersection of their capabilities, and that means no uses of reserve or capacity (see Item 14), because deque and list don't offer them. The presence of list also means you give up operator[], and you limit yourself to the capabilities of bidirectional iterators. That, in turn, means you must stay away from algorithms that demand random access iterators, including sort, stable_sort, partial_sort, and nth_element (see Item 31).

On the other hand, your desire to support vector rules out use of push_front and pop_front, and both vector and deque put the kibosh on splice and the member form of sort. In conjunction with the constraints above, this latter prohibition means that there is no form of sort you can call on your "generalized sequence container."

That's the obvious stuff. If you violate any of those restrictions, your code will fail to compile with at least one of the containers you want to be able to use. The code that *will* compile is more insidious.

The main culprit is the different rules for invalidation of iterators, pointers, and references that apply to different sequence containers. To write code that will work correctly with vector, deque, and list, you must assume that any operation invalidating iterators, pointers, or references in any of those containers invalidates them in the container you're using. Thus, you must assume that every call to insert invalidates everything, because deque::insert invalidates all iterators and, lacking the ability to call capacity, vector::insert must be assumed to invalidate all pointers and references. (Item 1 explains that deque is unique in sometimes invalidating its iterators without invalidating its pointers and references.) Similar reasoning leads to the conclusion that, unless you're eraseing the last element of a container, calls to erase must also be assumed to invalidate everything.

Want more? You can't pass the data in the container to a C interface, because only vector supports that (see Item 16). You can't instantiate your container with bool as the type of objects to be stored, because, as Item 18 explains, vector<bool> doesn't always behave like a vector, and it never actually stores bools. You can't assume list's constant-time insertions and erasures, because vector and deque take linear time to perform those operations.

When all is said and done, you're left with a "generalized sequence container" where you can't call reserve, capacity, operator[], push_front, pop_front, splice, or any algorithm requiring random access iterators; a container where every call to insert and erase takes linear time and invalidates all iterators, pointers, and references; and a container incompatible with C where bools can't be stored. Is that really the kind of container you want to use in your applications? I suspect not.

If you rein in your ambition and decide you're willing to drop support for list, you still give up reserve, capacity, push_front, and pop_front; you still must assume that all calls to insert and erase take linear time and invalidate everything; you still lose layout compatibility with C; and you still can't store bools.

If you abandon the sequence containers and shoot instead for code that can work with different associative containers, the situation isn't much better. Writing for both set and map is close to impossible, because sets store single objects while maps store pairs of objects. Even writing for both set and multiset (or map and multimap) is tough. The insert member function taking only a value has different return types for sets/maps than for their multi cousins, and you must religiously avoid making any assumptions about how many copies of a value are stored in a container. With map and multimap, you must avoid using operator[], because that member function exists only for map.

Face the truth: it's not worth it. The different containers are *different*, and they have strengths and weaknesses that vary in significant ways. They're not designed to be interchangeable, and there's little you can do to paper that over. If you try, you're merely tempting fate, and fate doesn't like to be tempted.

Still, the day will dawn when you'll realize that a container choice you made was, er, suboptimal, and you'll need to use a different container type. You now know that when you change container types, you'll not only need to fix whatever problems your compilers diagnose, you'll also need to examine all the code using the container to see what needs to be changed in light of the new container's performance characteristics and rules for invalidation of iterators, pointers, and references. If you switch from a vector to something else, you'll also have to make sure you're no longer relying on vector's C-compatible memory layout, and if you switch to a vector, you'll have to ensure that you're not using it to store bools.

Given the inevitability of having to change container types from time to time, you can facilitate such changes in the usual manner: by encapsulating, encapsulating, encapsulating. One of the easiest ways to do this is through the liberal use of typedefs for container types. Hence, instead of writing this,

```
class Widget { ... };
vector<Widget> vw;
Widget bestWidget;
...                                          // give bestWidget a value
vector<Widget>::iterator i =                 // find a Widget with the
  find(vw.begin(), vw.end(), bestWidget);    // same value as bestWidget
```

write this:

```
class Widget { ... };
typedef vector<Widget> WidgetContainer;
WidgetContainer cw;
Widget bestWidget;
...
WidgetContainer::iterator i = find(cw.begin(), cw.end(), bestWidget);
```

This makes it a lot easier to change container types, something that's especially convenient if the change in question is simply to add a custom allocator. (Such a change doesn't affect the rules for iterator/pointer/reference invalidation.)

```
class Widget { ... };

template<typename T>                          // see Item 10 for why this
SpecialAllocator { ... };                     // needs to be a template

typedef vector<Widget, SpecialAllocator<Widget> > WidgetContainer;

WidgetContainer cw;                           // still works

Widget bestWidget;

...

WidgetContainer::iterator i =
  find(cw.begin(), cw.end(), bestWidget);     // still works
```

If the encapsulating aspects of typedefs mean nothing to you, you're still likely to appreciate the work they can save, especially for iterator types. For example, if you have an object of type

```
map< string,
     vector<Widget>::iterator,
     CIStringCompare>                    // CIStringCompare is "case-
                                         // insensitive string compare;"
                                         // Item 19 describes it
```

and you want to walk through the map using const_iterators, do you really want to spell out

```
map<string, vector<Widget>::iterator, CIStringCompare>::const_iterator
```

more than once? Once you've used the STL a little while, you'll realize that typedefs are your friends.

A typedef is just a synonym for some other type, so the encapsulation it affords is purely lexical. A typedef doesn't prevent a client from doing (or depending on) anything they couldn't already do (or depend on). You need bigger ammunition if you want to limit client exposure to the container choices you've made. You need classes.

To limit the code that may require modification if you replace one container type with another, hide the container in a class, and limit the amount of container-specific information visible through the class interface. For example, if you need to create a customer list, don't use a list directly. Instead, create a CustomerList class, and hide a list in its private section:

```
class CustomerList {
private:
  typedef list<Customer> CustomerContainer;
  typedef CustomerContainer::iterator CCIterator;

  CustomerContainer customers;

public:                                 // limit the amount of list-specific
  ...                                   // information visible through
};                                      // this interface
```

At first, this may seem silly. After all a customer list is a *list*, right? Well, maybe. Later you may discover that you don't need to insert or erase customers from the middle of the list as often as you'd anticipated, but you do need to quickly identify the top 20% of your customers — a task tailor-made for the nth_element algorithm (see Item 31). But nth_element requires random access iterators. It won't work with a list. In that case, your customer "list" might be better implemented as a vector or a deque.

When you consider this kind of change, you still have to check every CustomerList member function and every friend to see how they'll be affected (in terms of performance and iterator/pointer/reference invalidation, etc.), but if you've done a good job of encapsulating CustomerList's implementation details, the impact on CustomerList clients should be small. *You* can't write container-independent code, but *they* might be able to.

Item 3: Make copying cheap and correct for objects in containers.

Containers hold objects, but not the ones you give them. Instead, when you add an object to a container (via, e.g., insert or push_back, etc.), what goes into the container is a *copy* of the object you specify.

Once an object is in a container, it's not uncommon for it to be copied further. If you insert something into or erase something from a vector, string, or deque, existing container elements are typically moved (copied) around (see Items 5 and 14). If you use any of the sorting algorithms (see Item 31); next_permutation or previous_permutation; remove, unique, or their ilk (see Item 32); rotate or reverse, etc., objects will be moved (copied) around. Yes, copying objects is the STL way.

It may interest you to know how all this copying is accomplished. That's easy. An object is copied by using its copying member functions, in particular, its *copy* constructor and its *copy* assignment operator. (Clever names, no?) For a user-defined class like Widget, these functions are traditionally declared like this:

```
class Widget {
public:
  ...
  Widget(const Widget&);                    // copy constructor
  Widget& operator=(const Widget&);         // copy assignment operator
  ...
};
```

As always, if you don't declare these functions yourself, your compilers will declare them for you. Also as always, the copying of built-in types (e.g., ints, pointers, etc.) is accomplished by simply copying the underlying bits. (For details on copy constructors and assignment operators, consult any introductory book on C++. In *Effective C++*, Items 11 and 27 focus on the behavior of these functions.)

With all this copying taking place, the motivation for this Item should now be clear. If you fill a container with objects where copying is expensive, the simple act of putting the objects into the container could prove to be a performance bottleneck. The more things get moved around in the container, the more memory and cycles you'll blow on making copies. Furthermore, if you have objects where "copying" has an unconventional meaning, putting such objects into a container will invariably lead to grief. (For an example of the kind of grief it can lead to, see Item 8.)

In the presence of inheritance, of course, copying leads to slicing. That is, if you create a container of base class objects and you try to insert derived class objects into it, the derivedness of the objects will be removed as the objects are copied (via the base class copy constructor) into the container:

```
vector<Widget> vw;

class SpecialWidget:                 // SpecialWidget inherits from
  public Widget { ... };             // Widget above

SpecialWidget sw;

vw.push_back(sw);                    // sw is copied as a base class
                                     // object into vw. Its specialness
                                     // is lost during the copying
```

The slicing problem suggests that inserting a derived class object into a container of base class objects is almost always an error. If you want the resulting object to *act* like a derived class object, e.g., invoke derived class virtual functions, etc., it *is* always an error. (For more background on the slicing problem, consult *Effective C++*, Item 22. For another example of where it arises in the STL, see Item 38.)

An easy way to make copying efficient, correct, and immune to the slicing problem is to create containers of *pointers* instead of containers of objects. That is, instead of creating a container of Widget, create a container of Widget*. Copying pointers is fast, it always does exactly what you expect (it copies the bits making up the pointer), and nothing gets sliced when a pointer is copied. Unfortunately, containers of pointers have their own STL-related headaches. You can read about them in Items 7 and 33. As you seek to avoid those headaches while still dodging efficiency, correctness, and slicing concerns, you'll probably discover that containers of *smart pointers* are an attractive option. To learn more about this option, turn to Item 7.

If all this makes it sound like the STL is copy-crazy, think again. Yes, the STL makes lots of copies, but it's generally designed to avoid copying objects *unnecessarily*. In fact, it's generally designed to avoid *creating* objects unnecessarily. Contrast this with the behavior of C's and C++'s only built-in container, the lowly array:

```
Widget w[maxNumWidgets];   // create an array of maxNumWidgets
                           // Widgets, default-constructing each one
```

This constructs maxNumWidgets Widget objects, even if we normally expect to use only a few of them or we expect to immediately overwrite each default-constructed value with values we get from someplace else (e.g., a file). Using the STL instead of an array, we can use a vector that grows when it needs to:

```
vector<Widget> vw;            // create a vector with zero Widget
                              // objects that will expand as needed
```

We can also create an empty vector that contains enough space for maxNumWidgets Widgets, but where zero Widgets have been constructed:

```
vector<Widget> vw;
vw.reserve(maxNumWidgets);    // see Item 14 for details on reserve
```

Compared to arrays, STL containers are much more civilized. They create (by copying) only as many objects as you ask for, they do it only when you direct them to, and they use a default constructor only when you say they should. Yes, STL containers make copies, and yes, you need to understand that, but don't lose sight of the fact that they're still a big step up from arrays.

Item 4: Call empty instead of checking size() against zero.

For any container c, writing

```
if (c.size() == 0) ...
```

is essentially equivalent to writing

```
if (c.empty()) ...
```

That being the case, you might wonder why one construct should be preferred to the other, especially in view of the fact that empty is typically implemented as an inline function that simply returns whether size returns 0.

You should prefer the construct using empty, and the reason is simple: empty is a constant-time operation for all standard containers, but for some list implementations, size may take linear time.

But what makes list so troublesome? Why can't it, too, offer a constant-time size? The answer has much to do with the range form of list's unique splicing functions. Consider this code:

```
list<int> list1;
list<int> list2;

...

list1.splice(                                   // move all nodes in list2
  list1.end(), list2,                           // from the first occurrence
  find(list2.begin(), list2.end(), 5),          // of 5 through the last
  find(list2.rbegin(), list2.rend(), 10).base() // occurrence of 10 to the
);                                              // end of list1. See Item 28
                                                // for info on the "base()" call
```

This code won't work unless list2 contains a 10 somewhere beyond a 5, but let's assume that's not a problem. Instead, let's focus on this question: how many elements are in list1 after the splice? Clearly, list1 after the splice has as many elements as it did before the splice plus however many elements were spliced into it. But how many elements were spliced into it? As many as were in the range defined by find(list2.begin(), list2.end(), 5) and find(list2.rbegin(), list2.rend(), 10).base(). Okay, how many is that? Without traversing the range and counting them, there's no way to know. And therein lies the problem.

Suppose you're responsible for implementing list. list isn't just any container, it's a *standard* container, so you know your class will be widely used. You naturally want your implementation to be as efficient as possible. You figure that clients will commonly want to find out how many elements are in a list, so you'd like to make size a constant-time operation. You'd thus like to design list so it always knows how many elements it contains.

At the same time, you know that of all the standard containers, only list offers the ability to splice elements from one place to another without copying any data. You reason that many list clients will choose list specifically because it offers high-efficiency splicing. They know that splicing a range from one list to another can be accomplished in constant time, and you know that they know it, so you certainly want to meet their expectation that splice is a constant-time member function.

This puts you in a quandary. If size is to be a constant-time operation, each list member function must update the sizes of the lists on which it operates. That includes splice. But the only way for the range version of splice to update the sizes of the lists it modifies is for it to count the number of elements being spliced, and doing that would prevent it from achieving the constant-time performance you want for *it*. If you eliminate the requirement that the range form of splice update the sizes of the lists it's modifying, splice can be made constant-time, but then size becomes a linear-time operation. In general, it will have to traverse its entire data structure to see how many elements it contains. No matter how you look at it, something — size or the range form of splice — has to give. One or the other can be a constant-time operation, but not both.

Different list implementations resolve this conflict in different ways, depending on whether their authors choose to maximize the efficiency of size or the range form of splice. If you happen to be using a list implementation where a constant-time range form of splice was given higher priority than a constant-time size, you'll be better off calling empty than size, because empty is always a constant-time operation. Even if you're not using such an implementation, you might find yourself using such an implementation in the future. For example, you might port your code to a different platform where a different implementation of the STL is available, or you might just decide to switch to a different STL implementation for your current platform.

No matter what happens, you can't go wrong if you call empty instead of checking to see if size() == 0. So call empty whenever you need to know whether a container has zero elements.

Item 5: Prefer range member functions to their single-element counterparts.

Quick! Given two vectors, v1 and v2, what's the easiest way to make v1's contents be the same as the second half of v2's? Don't agonize

Index

The example classes and class templates declared or defined in this book are indexed under *example classes/templates*. The example functions and function templates are indexed under *example functions/templates*.

Before A

A

C

D

E

F

M

N

O

P

Q

R

S