Runtime Checking C Programs

Reed Milewicz
University of Alabama at Birmingham
Department of Computer and Information Sciences
Birmingham, AL 35294
rmmilewi@cis.uab.edu

Rajesh Vanka
Matlab
3 Apple Hill Dr
rvanka@ncsu.edu

James Tuck
North Carolina State University
Department of Electrical and Computer Engineering
jtuck@ncsu.edu

Daniel Quinlan
Lawrence Livermore National Laboratory
Livermore, CA 94550
dquinlan@llnl.gov

Peter Pirkelbauer
University of Alabama at Birmingham
Department of Computer and Information Sciences
Birmingham, AL 35294
pirkelbauer@uab.edu

ABSTRACT

The C Programming Language is known for being an efficient language that can be compiled on almost any architecture and operating system. However the absence of dynamic safety checks and a relatively weak type system allows programmer oversights that are hard to spot. In this paper, we present RTC, a runtime monitoring tool that instruments unsafe code and monitors the program execution. RTC is built on top of the ROSE compiler infrastructure. RTC finds memory bugs and arithmetic overflows and underflows, and run-time type violations. Most of the instrumentations are directly added to the source file and only require a minimal runtime system. As a result, the instrumented code remains portable. In tests against known error detection benchmarks, RTC found 98% of all memory related bugs and had zero false positives. In performance tests conducted with well known algorithms, such as binary search and MD5, we determined that the unoptimized overhead rate is between a factor of 1.8 and a factor of 77 respectively.

1. INTRODUCTION

One major trend in computing is the continuing increase in the complexity of software systems. Such an increase is allowed by the expectation of increasingly powerful hardware (faster processors, larger memory and disks) and the increasing diversity of environments in which the software runs. This increase in complexity is expensive; the National Institute for Standards and Technology (NIST) estimated that inadequate infrastructure for software testing costs the US economy $22.2 billion annually [26].

Many programming languages allow the use of unsafe programming constructs in order to attain a high degree of flexibility and performance. This makes the construction of correct large-scale software difficult. Unsafe language features allow programmer oversights to introduce software flaws which can create security hazards that can compromise an entire system. Eliminating these software flaws can be addressed on many levels in the software engineering process. Rigid coding standards, restricting the use of a large language to a safer subset, peer review, and the use of static or dynamic analysis tools are some means that can all reduce the exposure to software flaws.

Analysis tools can be categorized by how they analyze software. Static tools analyze software without running it. They target source code (occasionally binary) to apply formal analysis techniques such as dataflow analysis, abstract interpretation, and model checking; such techniques often use approximations to arrive at sound but imprecise conclusions about the behavior of programs. Dynamic analysis tools find bugs by observing the behavior of running programs. This is typically accomplished by code instrumentation (source or binary) or by replacing (built-in) library functions (e.g., malloc and free) with custom implementations. Dynamic analysis operates with concrete values and is not prone to combinatorial state explosion. The downside of dynamic analysis tools is that monitoring running programs impacts performance. The quality of the results depends on the tests' input data and covered program paths. Hybrid analyses combine static and dynamic techniques. Hybrid tools can improve performance by eliminating checks that a static analyzer considers safe in all possible scenarios, or they can improve test coverage by producing input values that cover all possible paths.

The history of program analysis research now spans several decades, and dozens of tools and techniques have pushed the envelope on our ability to detect and correct bugs. However, in the area of dynamic analysis, certain fundamental challenges remain, namely high overhead costs and lack of portability. Neither of these challenges have gone unnoted, but have been relatively low priority targets for researchers.

SAC '15 ,April 13 - 17 2015, Salamanca, Spain (c) 2014 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.
ACM 978-1-4503-3196-8/15/04 ...$15.00. http://dx.doi.org/10.1145/2578726.2578744

Severe overhead, in many respects, has been seen as the cost of doing business, a burden that software developers have been willing to tolerate. For dynamic analysis tools that target binaries for instrumentation (the most popular choice), high overhead costs are difficult to avoid. The most efficient way to reduce overhead is to selectively reduce instrumentation where bugs are unlikely or are rendered impossible, but without access to high-level information (e.g. type information) to guide the pruning process, this can be a risky proposition. Meanwhile, the computing landscape has been historically dominated by only a handful of operating systems and varieties of architecture; meeting the portability requirements of developers meant being able to operate in two or three popular environments.

However, as we move towards a future where computing pervades every aspect of our daily lives, these challenges become more substantial and more must be done to address them. The most visible signal of the shift to ubiquitous computing has been the proliferation of smartphones and the thriving ecosystem of services that interface with them. However, the most influential transformations have come from the progressive infiltration of embedded computing systems, from critical control devices in medical care and avionics, to smart televisions and coffee machines. Many of the most popular languages for development in these burgeoning environments are also the least safe (e.g. C and C++), the improper use of which introduce vulnerabilities that threaten reliability and security. At the same time, burdensome overhead costs and portability limitations render *in situ* dynamic analysis impractical if not impossible. The imperative is a simple one: adapt or die.

In this paper, we present RTC (Run-Time error check for C programs), a dynamic analysis tool for C99 programs. RTC instruments source code with safety checks and produces another C source file. The resulting source file is portable and can be compiled on any platform and any compiler that can handle C99 and linked to a small runtime system. Choosing to instrument source code instead of binary code has a number of advantages. First, the tool is portable, because the systems where the code is instrumented and the system where the code runs can be different architectures. The only requirement is that there is a C99 compiler available for the target system. Second, by instrumenting source code, the tool processes the code as written by the programmer, and not some code that was generated and possibly optimized by a compiler. Finally, we can choose to validate only a single program module. For example, we may want to only a single, commonly used library. In such cases, we may want to instrument only that library and not the whole application.

Currently, RTC supports C99 and a subset of sequential C++. RTC implements three kinds of safety checks: Arithmetic overflow/underflow, memory safety checks to find memory bugs on stack and heap, and run-time type-safety violations. The metadata is kept on the side using a locks and keys approach. Arithmetic overflow/underflow and memory safety checks cover three of the most dangerous software bugs [24]. We tested RTC on several runtime checking benchmarks and on complete programs including grep, crafty, and other C programs in the SPEC2000 benchmark suite.

The paper presents the following contributions: (1) automated and portable source code instrumentation and monitoring for C99 programs; (2) lightweight runtime monitoring implementation.

The paper is outlined as follows: §2 presents background information and related work. §3 describes our implementation in detail, and §4 discusses how we tested our tool and the obtained results. §5 summarizes the paper and discusses possible future research directions.

2. BACKGROUND

In this section, we present earlier work on error checking, and an overview of the ROSE source-to-source transformation system.

2.1 Related Work

Run-time error checking tools have been designed for a variety of reasons, ranging from bug detection and security to software verification.

SafeC [1] introduced the notion of using source code instrumentation to detect temporal and spatial memory errors. CCured [18], Cyclone [10], and MSCC [28] are all works derived from or inspired by SafeC. Of particular interest to us is CCured, which introduced the notion of using lightweight, disjoint metadata facilities to cut down on the massive overhead inherent in the approach taken by SafeC. This idea inspired Hardbound [5], which attempts to tackle the issue by giving hardware support for bounds checking pointers and pointer management. A software-based approach analogous to HardBound was explored in SoftBound [17]. MemSafe [25] extends this idea by using static analysis to prove memory accesses safe. ConSeq [29] identifies code locations, such as assertions or reads of key global variables. Then ConSeq extracts slices to determine instructions that contribute to violations. Finally, ConSeq uses dynamic analysis to generate valid, bug-free executions of the program, and then see whether any legal deviations from that execution could lead to the potential errors that were revealed by the static analysis.

RTED [23] is a dynamic analysis tool developed at the Lawrence Livermore National Laboratory. RTED is a first proof of concept implementation for sequential C, a large sequential subset of C++, and UPC[6], a parallel language for the partitioned global address space model. RTED finds temporal and spatial memory violations on stack and heap, signature mismatch in declaration and definition, erroneous library calls, and reads from uninitialized memory. RTED statically inserts source code to monitor the execution. For concurrent codes in UPC, RTED synchronizes shared memory accesses to force a deterministic execution. Consequently, the runtime overhead of RTED is large (>100x). For the targeted errors, the error detection tool achieved roughly 95% coverage on Iowa State's runtime error detection benchmark suite [13].

Frama-C [12] is an optimized runtime memory monitoring library that implements assertion checking, memory and pointer safety checking as library. Frama-C is complemented by E-ACSL which is a first-order logic annotation language. The specifications can be translated to Frama-C runtime checks.

Other verification tools are also available. Valgrind [20], a framework for writing dynamic binary analyses tools. Valgrind finds memory access violations, concurrency related bugs, and others. It supports concurrency, but serializes concurrent executions. IBM's Rationale Purify [9] is memory debugger that instruments object code and tracks memory allocation and initializations. Parasoft's Insure++ [22] is a proprietary tool that instruments source and monitors execution in threaded applications. CDSChecker [21] is a stateless model checker for concurrent software written in the C11 and C++11 programming languages.

Other parallel error checking tools are geared to find concurrency bugs [2, 8, 15]. Some runtime error checking approaches require support built into hardware [4, 7, 14, 16, 30]. Probabilistic methods allow to deal with uncertainty resulting from less frequent program observation [11].

Lastly, we note that [3] explored run-time type-checking in C, by using binary instrumentation informed by debugging information to perform fine-grained type-safety analysis.

2.2 The ROSE transformation system

ROSE, developed at the Lawrence Livermore National Laboratory (LLNL), is a source-to-source translation infrastructure for multiple languages, including C/C++, Fortran 77/95/2003, Java, and UPC. ROSE also supports several extensions to develop parallel programs, such as OpenMP and CUDA. ROSE represents source code as abstract syntax trees (AST). The ASTs are built in a uniform and consistent way for all input languages. ROSE implements many specific analyses (e.g. pointer alias analysis) and makes them available through an API. Users can write their own analysis by utilizing frameworks that ROSE provides. These include attribute evaluation traversals, call graph analysis, control flow graphs, class hierarchies, SSA representation, and dataflow analysis. ROSE has been used for building custom tools for static analysis, program optimization, arbitrary program transformation, domain-specific optimizations, performance analysis, and cyber-security. ROSE can regenerate `include` directives, which maintains the portability of original code.

3. IMPLEMENTATION

We shall attempt to provide an overview of the implementation of the RTC tool. First, RTC supplies the original C source code to the ROSE compiler framework, which parses the input to produce an internal intermediate representation (IR), including a detailed abstract syntax tree (AST). The IR is, in turn, provided to RTC. RTC makes a pass over the AST to identify regions of code that require monitoring (such as when memory is allocated or a pointer is returned from a function call). Our tool then instruments the code by decorating the AST. Finally, the AST is unparsed, yielding the instrumented C source code. This code can then be compiled, linked to RTC's metadata libraries, and then executed. The monitoring infrastructure put in place during the instrumentation phase allows us to probe the behavior of the program at runtime to detect bugs. A diagram illustrating this process can be seen in Fig. 1.

The following subsections provide detailed descriptions of each of the aforementioned steps.

3.1 Preprocessing

When RTC receives the AST from ROSE, the AST must be preprocessed. The purpose of the preprocessing phase is to make the AST more amenable to instrumentation while preserving the semantics of the program. By requiring that the AST be normalized first, we reduce the complexity of the instrumentation process and ensure that instrumentation is consistently and correctly applied. The preprocessing step is split up over many distinct functions, each of which queries the AST to find and perform normalizing operations on select nodes of the AST. These transformations include the following:

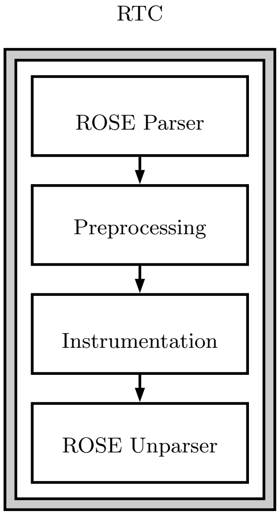

Figure 1: The RTC approach.

1. The termination conditions of for and while loops are moved inside their associated blocks. This guarantees that all for and while loops have similar structures.
2. Arrow expressions are converted to dot expressions. This allows us to treat arrow and dot expressions in the same fashion.
3. Structs defined within functions are moved out into the global scope. Otherwise, the instrumentation functions might overlook the struct definitions.

Once all has been laid bare, RTC performs a complete traversal of the AST, composing a list of every node that requires instrumentation. The resulting list is then passed to the instrumentation phase.

3.2 Instrumentation

The purpose of the instrumentation phase is to guarantee that every interaction with memory is guarded by appropriate calls to runtime monitoring functions, and that all pointers are outfitted with metadata that allows us to track when, where, and how they are used. At the same time, the panoptical ideal can only be realized within certain constraints: we must apply the instrumentation consistently and preserve the original behaviors of the program. The preprocessing phase guarantees the first requirement, but not the latter, which is the subject of this subsection. For the purposes of demonstrating this process, we have taken an abridged excerpt from version 1.0.1 of the OpenSSL library, containing the code responsible for the Heartbleed bug which was first disclosed in April 2014. The exploit relies on a spatial memory violation, and we shall demonstrate how RTC instruments this code to expose the violation at run-time.

The first objective of the instrumentation process is to expose all pointers so that their uses can be analyzed at runtime. For each pointer, RTC maintains a metadata record

```
void process_heartbeat(unsigned char *hbMessage) {
  unsigned short hbtype;
  unsigned int payload;
  unsigned char* contents;
  hbtype = hbMessage[0];
  payload = (((unsigned int)(hbMessage[1]))<< 8)
            | (((unsigned int)(hbMessage[2])));
  hbMessage += 3;
  contents = hbMessage;
  if (hbtype == TLS1_HB_REQUEST) {
    unsigned char* response;
    response = (unsigned char *) malloc(1 + 2 + payload + padding);
    ...
    memcpy(response, pl, payload);
    ...
  }
  ...
}
```

Figure 2: A simplified representation of the code responsible for the Heartbleed bug

that is kept in a separate and disjoint data structure. Pointers are moved into structures that contain two fields: one for the pointer and another to hold the stack address of that pointer; the latter serves as an index to that pointer's associated metadata. For every type of pointer in the input program, RTC declares and defines a struct to hold pointers of that type, as well as functions that handle the creation of those structs. In the case of our example, RTC provides structs for pointers to types `unsigned char` and `void`. An assignment to a pointer variable is split into a declaration and a call to a specialized assignment function that creates an instance of the struct and assigns the variable. Operations that alter pointers are substituted with calls to specialized functions that perform these operations and move the pointer into new wrapper structs; pointers are kept in constant motion, donning whatever raiment is appropriate for the present circumstance.

This system of pointer-wrapping structs also forms the backbone for the runtime type-checking system, as it provides the opportunity to record the data type of the pointer. This is done by mapping the stack address of the pointer to a type history, a list of types associated with the pointer, stored in chronological order. In practice, types are represented by special typeinfo nodes, which list essential details about the type, such as whether the type is a primitive or an aggregate, the `sizeof` the type, and the base type (if one exists). RTC determines what types are used in a program statically, and injects declarations for each typeinfo node into the global scope, as well as calls to metadata library functions to instantiate the typeinfo nodes at runtime. In this way, information about each type is computed only once. This is an instance of the flyweight design pattern: type histories consist of a sequence of pointers to these typeinfo nodes, instead of having many redundant copies. Type histories are retained until a pointer is deallocated. This type history can be analyzed on demand for runtime type safety violations, as we shall see in §3.

When a pointer is passed to a function, the associated stack address is pushed onto a globally accessible stack provided by the metadata libraries. Upon entering the function, the address is popped from the stack so that it and the pointer can be redressed in a new wrapper. By using an external stack, RTC avoids having to alter function signatures to make metadata information available across scopes. An important consequence of this is that it is possible to call instrumented code from uninstrumented code without having to make changes or adaptations due to syntactic mismatches - crucial for testing purposes.

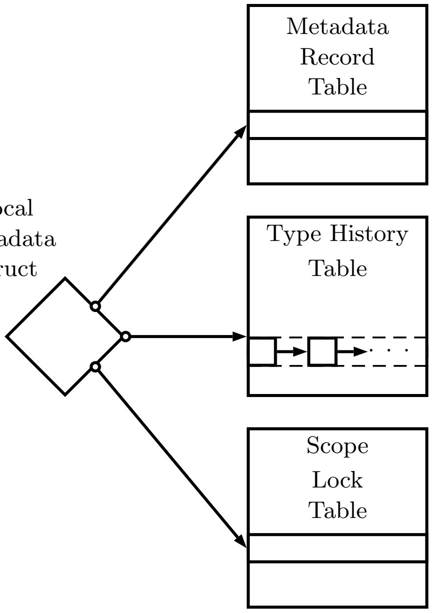

Figure 3: An overview of how RTC operates at runtime. Pointers are shadowed by lightweight metadata structures that hold the stack address of the pointer, used to index into the metadata record table, scope lock table, and type history table. Not shown is the metadata stack.

Once this has been done, RTC then inserts instrumentation around code where pointers are used to check for possible violations. As seen in Fig. 4, array accesses are guarded by calls to a bounds checking function that ensures that the access falls within the lower and upper bounds of the array allocation. Checks are also placed around calls to functions, `malloc`, `free`, and `pthread_create`, and other important library functions like `memcpy`. Calls to third party library functions that return pointers are replaced with calls to wrapper functions that handle the production of metadata information for the resulting pointers, as none is to be provided. The behavior of the original program is never altered by the addition of calls to the runtime monitoring library.

At the conclusion of the instrumentation phase, the AST is unparsed, giving us the instrumented source code file. After compiling and linking the output with the metadata libraries, we are left with an executable which can be run.

3.3 Runtime Checking

The runtime monitoring framework of RTC is divided into four components: the metadata reference stack, the metadata entry table, the scope lock table, and the type history table. The function of these structures, depicted in Fig. 3, is to help detect spatial and temporal errors in the use of memory in the instrumented program. To check for spatial errors, RTC uses the metadata gathered about the pointer to determine whether its corresponding location falls within

```
void process_heartbeat(unsigned char *hbMessage) {
  CREATE_MD_ENTRY_IF_EXISTS(&hbMessage,GET_FROM_STACK(0));
  unsigned short hbtype;
  unsigned int payload;
  unsigned char* contents;

  ARRAY_BOUNDS_CHECK(&hbMessage,&hbMessage[0]);
  hbtype = hbMessage[0];

  ARRAY_BOUNDS_CHECK(&hbMessage,&hbMessage[1]);
  ARRAY_BOUNDS_CHECK(&hbMessage,&hbMessage[2]);
  payload = (((unsigned int)(hbMessage[1]))<< 8)
              | (((unsigned int)(hbMessage[2])));
  hbMessage += 3;

  ASSIGN_AND_CREATE_MD_ENTRY(contents,hbMessage);

  if (hbtype == TLS1_HB_REQUEST) {
    unsigned char* response;
        ASSIGN_AND_CREATE_MD_ENTRY(response,
            CAST_UCHAR_PTR(
                MALLOC_WRAPPER(1 + 2 + payload + padding)));
    ...
        PUSH_TO_STACK(&response);
        PUSH_TO_STACK(&contents);
    MEMCPY_WRAPPER(response, contents, payload);
    ...
  }
  ...
}
```

Figure 4: The instrumented version of the `process_heartbeat` function from Fig. 2

the lower and upper bounds of the memory allocation. If it should stray, an error will result. Meanwhile, RTC uses a "lock and key" approach to detect temporal memory errors. Every scope in the program has a lock and key associated with it. When a pointer is brought into existence in a given scope, the values of the lock and key are written to that pointer's metadata entry, and a lock entry is created in a separate data structure that attests that the pointer is valid for use in that scope. When a pointer is freed, the lock entry is removed, and the lock/key values in the pointer's metadata are erased. To check for a temporal error in the use of a pointer, the key value held in the pointer's metadata is compared with the key of the current scope's lock. If there is a mismatch between the actual and expected key values, it means that the memory that is pointed to has been freed and/or reallocated, and an error is thrown. At the end of the execution of the program, RTC expects that all locks have been relinquished, which means that if a call to `malloc` is not matched with a corresponding call to `free` before the end of the program, an error will be reported; in this regard, RTC is more demanding of the programmer than the C memory model.

Finally, we shall describe the runtime type-checking system which RTC provides. As we noted previously, RTC is capable of collecting type histories of pointers that can be checked against to detect possible runtime type violations. One important function of this system is to catch uses of variables that have not been properly initialized, which can be determined by checking flags associated with the type history, a constant-time operation. Having a comprehensive type history also allows RTC to detect latent type violations, such as when non-`void` pointers are passed to functions as `void` pointers, only to be recast on the other side as a type that is incompatible with the original. To catch these errors, RTC can audit a pointer's type history, checking each type in sequence to ensure type-correctness. Auditing the type history is a linear-time operation, but in practice this does not significantly add to RTC's overhead; checking the whole type history of a pointer is a rare occurrence, and type histories tend to be short (usually no more than three or four entries).

As an object lesson in how RTC exercises these powers, we can look at the example program. The key issue is that the `process_heartbeat` function relies on an implicit assumption, that the length of the user's message, contained in `hbMessage` equals the reported length (stored in `payload`), and the programmer made no provisions to guard against the violation of that assumption. If a malicious client overstates the size of their message to the server (e.g. sending a 1 byte message while claiming the message is 64 kilobytes in size), the server's call to `memcpy` will fill the buffer `response` with the original 1 byte message, followed by 63980 bytes, taken from the contents of memory outside the bounds of `contents`.

When we enter the function `process_heartbeat`, RTC calls `CREATE_MD_ENTRY_IF_EXISTS` to check the stack address of the input string `hbMessage` against the address stored on the metadata stack. A metadata entry is created for the use of the `hbMessage` pointer in the current scope. Likewise, a separate metadata entry is created for the derived pointer `contents`. Finally, a metadata entry is created for the output buffer, `response`. When the program calls `MEMCPY_WRAPPER`, the wrapper function accesses the metadata associated with `contents` and `response`, to check that `payload` does not exceed the bounds of the allocation pointed to by `contents`. Then and only then can the call to `memcpy` proceed. If `memcpy` would read outside the bounds of `contents`, RTC raises an error and halts the execution.

4. EVALUATION

4.1 Estimating overhead costs

We collected a handful of C implementations of common algorithms to help give a sense of the overhead costs incurred by RTC. All implementations were single thread. The following is a list of the algorithms that were tested:

1. Binary Search: C implementation of the binary search algorithm. Searches a sorted array of elements in the hopes of finding a target value.
2. DTW: Dynamic time warping algorithm. Takes two input files containing numerical sequences and computes a measure of similarity between them.
3. Heapsort: Sort a list according to the heapsort algorithm. This is an in-place implementation.
4. MD5: A severely compromised cryptographic hash function.
5. Mergesort: Sort a list according to the mergesort algorithm.

Each of these algorithms was tested with a reasonable size of output and the execution times of the uninstrumented and instrumented implementations was compared. The experiments on an 8x AMD Opteron 6128 processor, running the Debian 6.0.5 operating system, and all code was compiled with version 4.4.5 of the GCC compiler with only default optimizations enabled. Each experiment was performed three times, and the average of these results were recorded. The

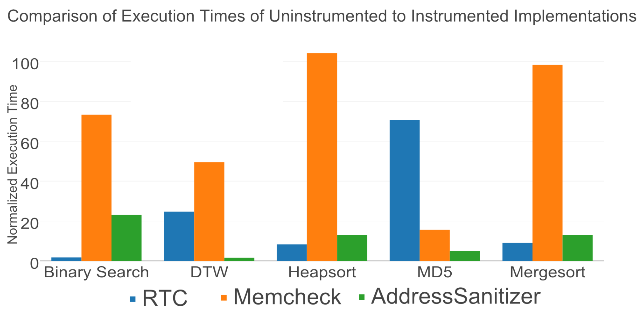

Figure 5: Normalized execution times for a selection of C implementations of algorithms instrumented by RTC, Clang's AddressSanitizer, and Valgrind's Memcheck.

results of these experiments can be seen in Fig. 5. For the sake of comparison, we also provide results for Clang's AddressSanitizer and Valgrind's Memcheck which we discuss in greater detail in subsection §4.3. These tools were chosen because they are popular, well-supported, and offer similar capabilities. We note that underflow and overflow detection was disabled for RTC and use-after-return detection was disabled for AddressSanitizer.

In the case of the binary search implementation, the bulk of the overhead comes from the bounds check added to every array access. For an input of N elements, binary search will require at most $\lfloor log_2(N) + 1 \rfloor$ iterations of the search, which means that there will also be that many calls to the array bounds checking function. A breakdown of the calls made to the metadata libraries reveals that these checks account for 97% of all calls. From this, we can estimate that these operations are responsible almost all of the 77% extra execution time in this case.

Next, comparing the results of the mergesort and heapsort implementations can show us the cost incurred by RTC's pointer monitoring facilities. Both mergesort and heapsort exhibit $O(NlogN)$ worst case performance, and we can expect the worst case cost in overhead for every additional element to be commensurate with this figure. As in the binary search implementation, both the mergesort and heapsort operations are dominated by array access operations in comparable measure, but that alone does not explain all of the overhead costs. We note that the mergesort implementation does use more memory than the heapsort implementation, but the only difference this makes in terms of the cost of the instrumentation is that an additional entry is made in the metadata table for every new temporary array, a relatively inexpensive operation. This is to say that the quantity of memory that is allocated by a program does not directly affect the overhead costs, only the extent to which that memory is used. However, we do note both implementations pass pointers to the functions they call, and all operations involving pointers are mediated by specialized structs and calls to the metadata library, and these can significantly add to the overhead costs.

This then brings us to the MD5 and DTW cases, which are representative of the kinds of real-world code that we are targeting with RTC. In both cases, the original source codes are awash with pointers, from calls to `malloc`, `realloc`, and `free`, to pointer dereferencing and type casting operations. Ensuring that all pointers are being used safely and correctly is no small task.

4.2 Code Coverage

To assess the bug coverage of RTC, we opted to test our tool against the Stonesoup test suite. Stonesoup refers to a curated test suite for C and Java published by the NIST in 2012, containing 460 cases dealing with issues of memory corruption and null pointer dereference. Each test case is well-documented, which means that we were able to instrument RTC to see whether it could detect the specific bug given in the description of the test case.

4.2.1 Palindrome Finder

One subcategory of the memory corruption test suite consists of variants of a program that takes a palindromic number as input and computes the next palindromic number. In each of these programs, a heap-based buffer overflow has been placed into the code that can occur when reading in the number to find the next palindrome. An attacker can provide input of a size larger than the intended maximum, and use this to launch a buffer overflow attack. Out of the 15 test cases, four of them used the X11 library to produce a GUI for the user to provide input, and these were excluded from consideration. Out of the remaining ten test cases, RTC succeeded at detecting bugs in nine. On the same test cases, Valgrind's Memcheck succeeded at detecting bugs in only five out of the ten. Clang's AddressSanitizer, meanwhile, aborted execution on six of the ten cases, having detected a segfault but being unable to explain why it occurred , and failed to find any bugs in the remaining four.

4.2.2 Solitaire Cipher

Another subcategory of the memory corruption test suite contains 20 variants of a solitaire encryption cipher program. In each case, the program was modified to include a buffer overflow vulnerability when reading in the key file that is used to create the random seed for shuffling. Five of them used the X11 library to handle user input, and these were excluded from consideration. Out the remaining 15 test cases, RTC succeeded in finding bugs in 6 of them. RTC failed to produce compilable code for the remaining cases, due to the a mishandling of an unusual array declaration. This indicates that more work is needed to ensure that RTC handles all of the corner cases of the C language. For the sake of comparison, Memcheck succeeded in finding bugs in 13 out of the 15 cases. AddressSanitizer aborted on 5 cases (for the same reason as in the palindrome tests), failed to detect any bugs on 6 cases, and successfully detected bugs on 4 cases.

4.3 Comparison of RTC to similar tools

Now we shall compare and contrast RTC to two other contemporary tools, Valgrind's Memcheck and Clang's AddressSanitizer. A summary of the points covered in this subsection can be found in Table 1.

4.3.1 Target of instrumentation

Both RTC and AddressSanitizer target source code for instrumentation; Memcheck targets binaries. The advantage for Memcheck is that it is completely language independent. However, working from the binary rather than the original source code means that type information is lost. One consequence is that invalid memory is reported when observable behavior is affected, but not when invalid memory is read. The lack of type information demands this behavioral definition, as compilers frequently emit load instructions for padding memory (e.g. a short value in a struct is aligned on word boundaries, and frequently compilers load the entire word instead of the smaller short). RTC and AddressSanitizer, meanwhile, can instrument all pointer accesses, which allows both tools to report uses of invalid memory when they occur.

4.3.2 *Use of shadow memory*

Both Memcheck and AddressSanitizer track shadow allocations of memory in order to determine when and where those allocations are valid for use. Valgrind's Memcheck tracks the validity of heap allocations on a per bit level, that is, one check bit for every bit of allocated memory. Clang's AddressSanitizer uses one bit of shadow memory per byte of real memory. With both tools, whenever memory is freed, that memory is marked as invalid for use and is rendered inacessible. In contrast, RTC shadows pointers, rather than allocations; the size of an allocation does not add to the amount of shadow memory required to track its use.

On a 64-bit architecture, a local metadata struct requires 128 bits of memory: 64 to hold the pointer itself and another 64 to hold the stack address of the pointer. The metadata entry, meanwhile, requires another 256 bits of memory: 128 for the lock and key associated with the scope of the pointer, and 128 to record both the lower and upper bounds of the allocation. Finally, an individual typetable entry consists of a variable-length sequence of 64-bit pointers to typeinfo nodes, and a pointer must have at least one type associated with it; this brings us to a minimum of 448 bits of information per pointer. Considering that the number of pointers and the size of allocations can vary from one program to the next, the relative size of RTC's shadow memory footprint can be difficult to estimate. However, on average, the number of pointers in use in a program at any one given time, multiplied by 448, is smaller than the quantity of allocated memory.

4.3.3 *Overhead*

On average, AddressSanitizer produces the lowest overhead of the three tools, followed by RTC and then by Memcheck. In practice, the actual overhead experienced depends heavily on the qualities of the program being instrumented. Because Memcheck requires very fine-grained monitoring, it is heavily affected by the size and quantity of memory allocations in addition to the volume of reads and writes from and to those allocations. AddressSanitizer employs a similar scheme but at one-eighth of the resolution, which, in conjunction with information derived from the source code (e.g. type information), reduces the costs associated with monitoring. Because RTC foregoes direct monitoring, its performance hinges upon the number of pointers used and the number of interactions with them, which we noted in our discussion of the findings seen on Fig. 5.

4.3.4 *Bug detection capabilities*

Unlike Memcheck but on par with AddressSanitizer, RTC is capable of tracking uses of memory on both the stack and the heap and has the ability to catch arithmetic underflow and overflow errors. Compared to AddressSanitizer, RTC offers the same features with the addition of support for run-time type-checking.

	Memcheck	AddressSanitizer	RTC
Target	Binary	Source Code	Source Code
Ratio of shadow memory to real memory	1:1	1:8	N/A
Average Slowdown	22x[19]	2x[27]	1.8x-77x
Out-of-bounds accesses (stack, heap, global)	No (heap only)	Yes	Yes
Use of uninitialized values	Yes	Yes	Yes
Invalid/double free	Yes	Yes	Yes
Use-after-free	Yes	Yes	Yes
Use-after-return	No	Yes	Yes
Memory Leaks	Yes	Yes	Yes
Arithmetic Overflows/ Underflows	No	Yes	Yes

Table 1: Comparison of RTC to other contemporary tools. Note that AddressSanitizer's support for use-after-return and memory leak detection are experimental, and use-after return detection is disabled by default.

5. CONCLUSION AND FUTURE WORK

In this paper, we have presented RTC, a source code instrumentation tool that finds software flaws common when using the C programming language. Our tool is capable of handling real-world programs. In our tests, we have demonstrated that RTC finds most bugs in available error detection benchmarks suits. As RTC is built with a source-to-source translation framework, we will integrate with static analyzers that are capable of proving many of the checks safe, allowing us to remove unnecessary safety checks. Our main research direction will extend our tool towards concurrency as permitted by the C11 and C14 programming languages.

6. REFERENCES

[1] Austin, T.M., Breach, S.E., Sohi, G.S.: Efficient detection of all pointer and array access errors. SIGPLAN Not. 29(6), 290–301 (jun 1994)

[2] Burnim, J., Elmas, T., Necula, G., Sen, K.: NDSeq: runtime checking for nondeterministic sequential specifications of parallel correctness. In: PLDI '11: Proceedings of the 32nd ACM SIGPLAN conference on Programming language design and implementation. ACM (Jun 2011)

[3] Burrows, M., Freund, S.N., Wiener, J.L.: Run-time type checking for binary programs. In: Proceedings of the 12th International Conference on Compiler Construction. pp. 90–105. CC'03, Springer-Verlag, Berlin, Heidelberg (2003)

[4] Chen, S., Kozuch, M., Strigkos, T., Falsafi, B., Gibbons, P.B., Mowry, T.C., Ramachandran, V., Ruwase, O., Ryan, M., Vlachos, E.: Flexible Hardware Acceleration for Instruction-Grain Program

Monitoring. Computer Architecture, 2008. ISCA '08. 35th International Symposium on pp. 377–388 (2008)

[5] Devietti, J., Blundell, C., Martin, M.M.K., Zdancewic, S.: Hardbound: Architectural support for spatial safety of the c programming language. In: Proceedings of the 13th International Conference on Architectural Support for Programming Languages and Operating Systems. pp. 103–114. ASPLOS XIII, ACM, New York, NY, USA (2008)

[6] El-Ghazawi, T., Carlson, W., Sterling, T., Yelick, K.: UPC Distributed Shared Memory Programming. Wiley Series on Parallel and Distributed Computing, Wiley, 1st edn. (2003)

[7] Falsafi, B., Gibbons, P.B., Kozuch, M., Mowry, T.C.: Log-based architectures for general-purpose monitoring of deployed code. In: Proceedings of the 1st Workshop on Architectural and System Support for Improving Software Dependability (2006)

[8] Goodstein, M.L., Vlachos, E., Chen, S., Gibbons, P.B., Kozuch, M.A., Mowry, T.C.: Butterfly analysis: adapting dataflow analysis to dynamic parallel monitoring. In: ASPLOS XV: Proceedings of the fifteenth edition of ASPLOS on Architectural support for programming languages and operating systems. ACM (Mar 2010)

[9] IBM: Rational PurifyPlus family. http://ibm.com/software/products/en/purifyplus/ (2014), accessed on March 12, 2014

[10] Jim, T., Morrisett, J.G., Grossman, D., Hicks, M.W., Cheney, J., Wang, Y.: Cyclone: A safe dialect of c. In: Proceedings of the General Track of the Annual Conference on USENIX Annual Technical Conference. pp. 275–288. ATEC '02, USENIX Association, Berkeley, CA, USA (2002)

[11] Kalajdzic, K., Bartocci, E., Smolka, S.A., Stoller, S.D., Grosu, R.: Runtime verification with particle filtering. In: Legay, A., Bensalem, S. (eds.) 4th International Conference on Runtime Verification (RV'13). Lecture Notes in Computer Science, vol. 8174, pp. 149–166. Springer Berlin Heidelberg (2013)

[12] Kosmatov, N., Petiot, G., Signoles, J.: An optimized memory monitoring for runtime assertion checking of c programs. In: Legay, A., Bensalem, S. (eds.) 4th International Conference on Runtime Verification (RV'13), Lecture Notes in Computer Science, vol. 8174, pp. 167–182. Springer Berlin Heidelberg (2013)

[13] Luecke, G.R., Coyle, J., Hoekstra, J., Kraeva, M., Xu, Y., Park, M.Y., Kleiman, E., Weiss, O., Wehe, A., Yahya, M.: The importance of run-time error detection. In: Parallel Tools Workshop. pp. 145–155 (2009)

[14] Mekkat, V., Holey, A., Zhai, A.: Accelerating data race detection utilizing on-chip data-parallel cores. In: Legay, A., Bensalem, S. (eds.) 4th International Conference on Runtime Verification (RV'13). Lecture Notes in Computer Science, vol. 8174, pp. 201–218. Springer Berlin Heidelberg (2013)

[15] Muzahid, A., Gracia, D.S., Qi, S., Torrellas, J.: SigRace: signature-based data race detection. ISCA pp. 337–348 (2009)

[16] Nagarakatte, S., Martin, M.M.K., Zdancewic, S.: WatchdogLite: Hardware-Accelerated Compiler-Based Pointer Checking. In: CGO '14: Proceedings of Annual IEEE/ACM International Symposium on Code Generation and Optimization. ACM (Feb 2014)

[17] Nagarakatte, S., Zhao, J., Martin, M.M.K., Zdancewic, S.: Softbound: Highly compatible and complete spatial memory safety for c. In: Proceedings of the 2009 ACM SIGPLAN Conference on Programming Language Design and Implementation. pp. 245–258. PLDI '09, ACM, New York, NY, USA (2009)

[18] Necula, G.C., Condit, J., Harren, M., McPeak, S., Weimer, W.: Ccured: Type-safe retrofitting of legacy software. ACM Trans. Program. Lang. Syst. 27(3), 477–526 (may 2005)

[19] Nethercote, N., Seward, J.: How to shadow every byte of memory used by a program. In: Proceedings of the 3rd International Conference on Virtual Execution Environments. pp. 65–74. VEE '07, ACM, New York, NY, USA (2007)

[20] Nethercote, N., Seward, J.: Valgrind: A framework for heavyweight dynamic binary instrumentation. SIGPLAN Not. 42(6), 89–100 (jun 2007)

[21] Norris, B., Demsky, B.: CDSchecker: Checking concurrent data structures written with C/C++ atomics. SIGPLAN Not. 48(10), 131–150 (oct 2013)

[22] Parasoft Inc.: Insure++. http://www.parasoft.com/insure (2014), accessed on March 12, 2014

[23] Pirkelbauer, P., Liao, C., Panas, T., Quinlan, D.: Runtime detection of C-style errors in UPC code. In: 5th Conference on Partitioned Global Address Space Models (PGAS). Galveston, TX (2011)

[24] SANS Institute: CWE/SANS TOP 25 most dangerous software errors (2011), http://www.sans.org/top25-software-errors

[25] Simpson, M.S., Barua, R.K.: Memsafe: ensuring the spatial and temporal memory safety of c at runtime. Software: Practice and Experience 43(1), 93–128 (2013)

[26] Tassey, G.: The Economic Impacts of Inadequate Infrastructure for Software Testing. NIST Report 02-3 (2002)

[27] The Clang Team: (sep 2014), `http://clang.llvm.org/docs/AddressSanitizer.html`

[28] Xu, W., DuVarney, D.C., Sekar, R.: An efficient and backwards-compatible transformation to ensure memory safety of c programs. In: Proceedings of the 12th ACM SIGSOFT Twelfth International Symposium on Foundations of Software Engineering. pp. 117–126. SIGSOFT '04/FSE-12, ACM, New York, NY, USA (2004)

[29] Zhang, W., Lim, J., Olichandran, R., Scherpelz, J., Jin, G., Lu, S., Reps, T.: Conseq: Detecting concurrency bugs through sequential errors. In: Proceedings of the Sixteenth International Conference on Architectural Support for Programming Languages and Operating Systems. pp. 251–264. ASPLOS XVI, ACM, New York, NY, USA (2011)

[30] Zhou, P., Qin, F., Liu, W., Zhou, Y., Torrellas, J.: iWatcher: efficient architectural support for software debugging. In: Computer Architecture, 2004. Proceedings. 31st Annual International Symposium on. pp. 224–235 (2004)