An Examination of the Usefulness of Language Extensions in an IRC Server

Will Trussell

Submitted under the supervision of Eric Van Wyk to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Science, *summa cum laude* in Computer Science.

20 November 2022

Abstract

Extensible languages, first introduced in the 1960s by McIlroy, offer a way of adding new syntax and semantics to a base language. This work utilizes an extensible version of C to develop and implement new extensions to improve and modify an implementation of an Internet Relay Chat (IRC) server. Three extensions are utilized in order to show that extensions can be useful and effective in simplifying the process of developing code used in modern software applications. Two of these three extensions introduce syntax to allow for compile-time checking of features that are not able to be checked at compile-time in C, while the third introduces new syntax to provide programmers with easier and more concise syntax for performing asynchronous I/O. Of the three extensions, two were written specifically for this work, while the third extension used was already written and just implemented for this work. This work not only examines the benefits provided by these extensions, but also includes an examination of the drawbacks of these extensions, primarily from the extension allowing for asynchronous I/O. This work demonstrates that the benefits of extensible programming are real and extensions, particularly those including compile-time checking of concepts not available in vanilla C, offer significant improvements in code writing and readability over the same software written in plain C.

Contents

1 Introduction

Extensible programming languages were first discussed in the 1960s. As originally discussed in Douglas McIlroy's 1960 paper, the original concept was to use a small number of macros in order to extend a compiler to accept very general extensions to the original language accepted by the compiler [**6**]. This has been modernized in various ways, including the idea of extensible syntax and compilers.

One example of an extensible programming language is `ABLE`C [**8**], an extensible version of the C programming language, built on top of the `Silver` [**3**] attribute grammar system. Various extensions exist for `ABLE`C that were created to solve various problems or add features. Such extensions include facilities for greater ease of writing parallel code in C, using syntax similar to Cilk[**2**], an implementation of closures in C, and the addition of algebraic data types and pattern matching.

This paper examines the uses of extensible programming languages in a real-world application: an Internet Relay Chat (IRC) server implementation. The IRC protocol was originally designed for text-based conferencing [**7**], and many server implementations are available. For this paper, we consider the specific implementation of `ngIRCd` (next-generation IRC daemon), written in the C programming language [**1**]. There are several potential issues with this server implementation. First, any C program must include runtime checks to ensure that all dereferenced pointers are not null. We introduce a pre-existing `ABLE`C extension to allow for both compile-time and runtime checking for possible null pointer dereferences. Another issue with the server is writing easily understandable I/O code in C. While the `epoll` API provided by the Linux kernel allows for some asynchronous I/O, we introduce a new `ABLE`C extension to provide improved syntax for asynchronous I/O, akin to what Javascript provides. Finally, we introduce a new extension to ensure that parsing code is secure from buffer overflow attacks. We do this utilizing the Wuffs programming language [**5**], which allows for compile-time checking for various parsing attacks which must be manually accounted for at run-time in C.

2 Background

This chapter provides background information for the work that follows throughout the paper. We first discuss the Silver attribute grammar system [**3**], which is used to implement the extensions to the host language. We follow this with a discussion of `ABLE`C [**8**], the extensible version of C utilized in this work.

2.1 Silver

Silver [**3**] is an extensible attribute grammar specification system that allows us to add both general features (pattern matching, for instance) and domain-specific features to the host language. This gives us an attribute grammar specification system with a rich set of language features that we utilize in defining the syntax and semantics of any new languages or extensions we may write.

Silver has several nice features useful in generating new language extensions. First and foremost, Silver allows for *forwarding* [**4**] to implement new extensions in cost-effective ways. Forwarding allows language designers to translate new syntax into the original host language. Using forwarding, language designers no longer need to explicitly define semantic analysis for each new piece of syntax in a language extension. Instead, any semantic analyses not defined by the language designer are instead defined via translation. A brief example of forwarding is given here; for a more in-depth discussion, see the examples given in Section 1 of the paper *Forwarding in Attribute Grammars for Modular Language Design* [**4**].

As an example, consider that we have some language with a defined string concatenation operator ++. Now consider adding an extension to the language, the operator $++_S$. We desire this operator to do the following: given the expression $s_1 + +_S s_2$, we would like the resulting expression to be s_2 *sandwiched* between s_1. Thus, we would like the result of $s_1 + +_S s_2$ to be $s_1 s_2 s_1$. We could fully define the semantics of the new $++_S$ operator, but this would be a waste of both time and effort. Instead, we can simply define our new operator

as follows.

```
1 sandwich : str_0 ::= str_1 str_2
2         forwards to concat(str_1, concat(str_2, str_1))
```

By doing this, we are able to quickly define a new piece of syntax for our language, allowing us to add new syntax and semantics to a language quickly and easily.

2.2 Extensible Programming and Able-C

One of the primary programming languages that is utilized in modern computing when speed or low-level control is vitally important is the C programming language. Unfortunately, C lacks many of the features of more modern programming languages, often making it cumbersome to work with in certain applications.

ABLEC [**9**] is a project that tries to improve upon the C language is through the use of extensions. Built utilizing Silver, ABLEC is an extensible C pre-processor, conforming to the C11 standard [**8**]. It takes an "extended" version of C and translates it back into plain C, performing transformations and analyses as it does so. For instance, there is an existing ABLEC extension that enables interaction with SQLite databases. With this extension, we begin with the ABLEC files. These files are then translated into plain C code after going through type-checking and error-checking by the ABLEC compiler. Another extension that provides a good example of the various analyses done by ABLEC is the ABLEC-refcount-closure extension. This extension implements reference counting of closures in C. The function performs analysis on how many references exist to a given closure to help with memory management. This is not a feature that is available in plain C code, but it is a useful tool for programmers to have in ABLEC.

3 Non-Null Pointer Extension

3.1 Non-null Qualifiers

One of the most common sources of bugs in the C programming language is errors related to null pointers. The dereferencing of a null pointer is considered undefined behavior in C, and is undesirable when programming. In this section, we introduce an ABLEC extension to deal with null pointers (referred to herein as ABLEC-nonnull). This extension allows for compile-time checking for any possible null dereferences of pointers. This static analysis of a program is valuable, saving the programmer both time and effort when compared with the original C code, which contains only dynamic checks at runtime.

3.2 Comparison

Consider the code in Figures 1 and 2. Both code snippets are from the same function in the `array.c` (or `array.xc`) file, with Figure 2 having been rewritten using ABLEC-nonnull. Note the differences between the two figures.

```
1  /* return false on failure (realloc failure, invalid src/dest array) */
2  bool
3  array_copyb(array * dest, const char *src, size_t len)
4  {
5          assert(dest != NULL);
6          assert(src != NULL );
7
8          if (!src || !dest)
9                  return false;
10
11         array_trunc(dest);
12         return array_catb(dest, src, len);
13 }
```

Figure 1: Function before applying the non-null extension

The code in Figure 1 is a simple function to copy a string into an array as it is written in the original project. It must explicitly check for null pointer dereferences using the `assert`

```
1 /* return false on failure (realloc failure, invalid src/dest array) */
2 bool
3 array_copyb(array * nonnull dest, const char * nonnull src, size_t len)
4 {
5
6         array_trunc(dest);
7         return array_catb(dest, src, len);
8 }
```

Figure 2: Function after applying the non-null extension

statements. Note that if this function is passed a null pointer for either of the two pointer arguments, the `assert` statement will fail, causing a call to the `abort()` function, crashing our program.

Compare this to Figure 2, which is the same function, but with the non-null extension. Here, we know that both pointer arguments are guaranteed to be non-null. Thus, we no longer need either of the assert statements, nor do we need the check in the `if` statement. This is a toy example implemented to illustrate the point that including this extension allows for shorter functions and fewer safety checks on the part of the end programmer.

As a final example of how this non-null extension works, we examine the translation of another function within the `array.xc` file. The purpose of this function is to append a trailing null byte to an array, but not count that byte in the length of the array. Note that in Figure 3, we are casting the results of a call to `array_alloc`. This is something the non-null extension will check at compile time. When we examine Figure 4, we see that the same cast is present, but there is a runtime check inserted after the cast to ensure that the pointer is not null. This is one way the extension guarantees that pointer dereferences are non-null.

```
1  /* append trailing NUL byte to array, but do not count it. */
2  bool array_cat0_temporary(array * nonnull a) {
3    * nonnull endpos = (char * nonnull) array_alloc(a, 1, array_bytes(a));
4    if (!endpos)
5      return false;
6    *endpos = '\0';
7    return true;
8  }
```

Figure 3: A function to append a null byte to an array

```
1  _Bool  array_cat0_temporary(array  * a) {
2    ;
3    {
4      char  *endpos = ((char *)({
5        void  *_tmp10 = ((array_alloc)((a), 1, ((((a))->used))));
6        ({
7          if (((_tmp10) == 0)) {
8            ((fprintf)((stderr),
9              "array.xc:194:25:ERROR: attempted cast of NULL to nonnull\\n"));
10           ((exit)(255));
11         } else {
12           ;
13         }
14         (_tmp10); })
15         ; })
16       );
17       if ((!(endpos)))
18       {
19         return 0;
20       } else {
21         ;
22       }
23       ((*(endpos)) = '\0');
24       return 1;
25     }
26 }
```

Figure 4: The same function to append a null byte to an array after being translated to C

4 Asynchronous I/O

4.1 Asynchronous I/O

Asynchronous I/O is a major part of the difficulty in creating a server implemenation in C, particularly when using threads is not a viable option. We don't want the server to be stuck while we wait for it to perform some I/O operation like reading or writing to an existing connection when the server has other tasks it could be doing, like establishing new client connections or parsing an already-received message. In C, one of the primary ways of performing asynchronous I/O is using the `epoll` API. Using this API, we are able to keep a list of file descriptors we want the current process to monitor, as well as a list of file descriptors that are ready for I/O. However, this process is extraordinarily tedious, requiring many expensive system calls to set up and maintain the `epoll` instance. When compared with other, more modern languages, the `epoll` API is both more verbose and more difficult to use. Consider a more modern language like Go or Javascript. Both of these languages have their own facilities for asynchronous operation. In Go, we use several constructs, including the `go` and `select` keywords, to implement various aspects of I/O. In Javascript, we utilize both the `async` and `await` keywords, as well as the idea of *promises* in order to achieve some measure of asynchronous operation. The goal of our extension is to include these same facilities in `ABLE`C, allowing for a programmer to more easily write and understand the code performing asynchronous I/O operations.

4.2 Asynchronous I/O Concepts

In this section, we discuss the mechanics of the Asynchronous I/O extension, including the specifics of the translation from `ABLE`C code to plain C code.

This extension is less focused on redesigning the fundamentals of asynchronous I/O and is more focused on providing syntax that is both easy to understand and easy to write for a programmer utilizing the extension. To this end, we retain some of the same principles

utilized by the `epoll` API, but introduce similar syntax to modern languages with the `spawn` and `await` keywords.

The `spawn` keyword has similar syntax to the `spawn` keyword in Cilk [**2**], the `spawn()` method in Ruby, or the `spawn()` method in Rust. All of these keywords have a similar idea that underlies them. Each of these languages uses `spawn` to indicate starting a new process, thread or function call. The syntax for all three is also similar in concept, as well. Each has the form `spawn <foo>`, where what is in `foo` is either a function or a closure (in Rust). This starts program execution on whatever task is passed through `foo`.

The `await` keyword is at least partially inspired by the syntax of Javascript. After we spawn several tasks using `spawn`, we can then specify that we would like to wait for those tasks using `await`. The syntax is similar to Javascript: we simply write `await <foo>`, where `foo`, after the `await` call, refers to a task that was created and run using the `spawn` keyword. In other words, `await` simply waits until at least one of the tasks we have spawned finishes, then loads that task (or tasks) into `foo`.

4.3 Asynchronous I/O Extension

Here we consider what changes are actually made by our extension. Consider the code snippets from the `io.c` file, shown below in Figure 5 and the code from the extension in Figure 6. The major changes between the figures are on lines 23, 45, and 50.

Note in the first example that we utilize the default `io_event_add()` and `io_dispatch()` functions. A particular issue that arises when dealing with these functions is the lack of high-level transparency in their calls. Perhaps `io_event_add()` is somewhat clear, but the `io_dispatch()` function is particularly opaque. We make a call to the function with some `timeval`, but we cannot determine the use of this function without substantial effort on the part of the programmer.

On the other hand, consider the implementation using the asynchronous extension. We are using the `spawn` keyword to create new tasks, in this case, calling out to a `read_helper()`

or `write_helper()` function. These functions are helper functions that simply perform a write or a read to some socket file descriptor, and we are passing them to `spawn` as the function to be run in a task. Note that because of these helper functions, we do not require any knowledge of constants like `IO_WANTREAD` or `IO_WANTWRITE`, instead simply requiring the programmer to pass in whatever function the programmer desires to run asynchronously.

This extension is also a much more flexible implementation than what is available with the `epoll` API. As an example, consider if a programmer wanted to modify the server to do something other than just read or write data using the `epoll` API. In the old C code, this would require defining a set of new constants and substantial code modification. In our new extension, all the programmer must be concerned with is writing whatever function they desire to be executed asynchronously. Then, utilize `spawn` with that function to register that function for asynchronous execution. There is no need to define any new constants or for the end programmer to do any substantial coding (outside of the function they desire to run asynchronously, which would have been written anyways without the extension).

```
1  GLOBAL void Conn_Handler(void) {
2          int i;
3          size_t wdatalen;
4          struct timeval tv;
5          time_t t;
6          bool command_available;
7          while (!NGIRCd_SignalQuit && !NGIRCd_SignalRestart) {
8                  t = time(NULL);
9                  command_available = false;
10                 //Utility checks omitted for brevity
11                 for (i = 0; i < Pool_Size; i++) { // Look for non-empty read buffers
12                         if ((My_Connections[i].sock > NONE)
13                             && (array_bytes(&My_Connections[i].rbuf) > 0)) {
14                                 Handle_Buffer(i); // handle the received data
15                         }
16                 }
17                 for (i = 0; i < Pool_Size; i++) { // Look for non-empty write buffers
18                         if (My_Connections[i].sock <= NONE)
19                                 continue;
20                         wdatalen = array_bytes(&My_Connections[i].wbuf);
21                         if (wdatalen > 0) {
22                                 //SSL Code omitted for brevity
23                                 io_event_add(My_Connections[i].sock, IO_WANTWRITE);
24                         }
25                 }
26                 for (i = 0; i < Pool_Size; i++) { //Check sockets for readability
27                         if (My_Connections[i].sock <= NONE)
28                                 continue;
29                         //SSL code omitted for brevity
30                         if (Proc_InProgress(&My_Connections[i].proc_stat)) {
31                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);//Wait on subprocesses
32                                 continue;
33                         }
34                         if (Conn_OPTION_ISSET(&My_Connections[i], CONN_ISCONNECTING))
35                                 continue; //Wait for connect() to complete
36                         if (My_Connections[i].delaytime > t) { //penalty set, ignore socket
37                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);
38                                 continue;
39                         }
40                         if (array_bytes(&My_Connections[i].rbuf) >= COMMAND_LEN) {
41                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);
42                                 command_available = true;
43                                 continue;
44                         }
45                         io_event_add(My_Connections[i].sock, IO_WANTREAD);
46                 }
47                 tv.tv_usec = 0;
48                 tv.tv_sec = command_available ? 0 : 1;
49                 i = io_dispatch(&tv);
50                 if (i == -1 && errno != EINTR) { exit(1); } //fatal errors
51                 if (Conf_IdleTimeout > 0 && NumConnectionsAccepted > 0
52                     && idle_t > 0 && time(NULL) - idle_t >= Conf_IdleTimeout) {
53                         NGIRCd_SignalQuit = true;
54                 }
55         }
56         //Server shutdown messages omitted
57 } /* Conn_Handler */
```

Figure 5: A file utilizing the asynchronous I/O Interface before implementing the extension

```
1  GLOBAL void Conn_Handler(void) {
2          int i;
3          size_t wdatalen;
4          struct timeval tv;
5          time_t t;
6          bool command_available;
7          while (!NGIRCd_SignalQuit && !NGIRCd_SignalRestart) {
8                  t = time(NULL);
9                  command_available = false;
10                 //Utility checks omitted for brevity
11                 for (i = 0; i < Pool_Size; i++) { // Look for non-empty read buffers
12                         if ((My_Connections[i].sock > NONE)
13                             && (array_bytes(&My_Connections[i].rbuf) > 0)) {
14                                 Handle_Buffer(i); // handle the received data
15                         }
16                 }
17                 for (i = 0; i < Pool_Size; i++) { // Look for non-empty write buffers
18                         if (My_Connections[i].sock <= NONE)
19                                 continue;
20                         wdatalen = array_bytes(&My_Connections[i].wbuf);
21                         if (wdatalen > 0) {
22                                 //SSL Code omitted for brevity
23                                 spawn write_helper(My_Connections[i].sock)
24                         }
25                 }
26                 for (i = 0; i < Pool_Size; i++) { //Check sockets for readability
27                         if (My_Connections[i].sock <= NONE)
28                                 continue;
29                         //SSL Code omitted for brevity
30                         if (Proc_InProgress(&My_Connections[i].proc_stat)) {
31                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);//Wait on subprocesses
32                                 continue;
33                         }
34                         if (Conn_OPTION_ISSET(&My_Connections[i], CONN_ISCONNECTING))
35                                 continue; //Wait for connect() to complete
36                         if (My_Connections[i].delaytime > t) { //penalty set, ignore socket
37                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);
38                                 continue;
39                         }
40                         if (array_bytes(&My_Connections[i].rbuf) >= COMMAND_LEN) {
41                                 io_event_del(My_Connections[i].sock, IO_WANTREAD);
42                                 command_available = true;
43                                 continue;
44                         }
45                         spawn read_helper(My_Connections[i].sock);
46                 }
47                 tv.tv_usec = 0;
48                 tv.tv_sec = command_available ? 0 : 1;
49                 io_event *events;
50                 i = await events; //Waits for some events, fills in events* with the events
51                 if (i == -1 && errno != EINTR) { exit(1); } //fatal errors
52                 if (Conf_IdleTimeout > 0 && NumConnectionsAccepted > 0
53                     && idle_t > 0 && time(NULL) - idle_t >= Conf_IdleTimeout) {
54                         NGIRCd_SignalQuit = true;
55                 }
56         }
57         //Server shutdown messages omitted
58 } /* Conn_Handler */
```

Figure 6: A file utilizing the asynchronous I/O interface after implementing the extension

5 Wuffs Parsing

5.1 Wuffs Overview

A final extension implemented for this project involves the Wuffs language. Wuffs is a language created by Google for memory-safe handling of untrusted file formats. Originally intended for use in security-specific portions of a program, Wuffs allows for compile-time checking of various security vulnerabilities, as well as some proofs of program safety.

A simple parsing program in C is shown below in Figure 7. The same parsing program is also shown in Figure 8, this time written in Wuffs as a comparison of the two languages. Both of these programs take as input a string of numbers, for example, `"123"`, and return the associated numeric value, in this case, 123.

Leaving aside the syntactic peculiarities of Wuffs, note that the Wuffs code is signficantly more verbose than the C code in the above examples. However, we consider the edge cases of parsing some string, particularly with regards to integer overflow. In C, an integer overflow error can occur, but it does so silently. Obviously, in a real C parser implementation, we would implement security checks to cover these overflow cases. The point, however, is less about the fact that this is possible in C and more about the fact that these checks are mandatory in Wuffs. The Wuffs compiler will not allow an addition expression that could possibly overflow the maximum integer value.

This compile-time security checking for basic security vulnerabilites is incredibly useful, particularly in contexts where we are parsing unknown code or strings that could be malicious. We don't need to worry about runtime security checks when we can use Wuffs to guarantee no buffer overflows, out-of-bounds array accessing, integer overflow, or integer underflow at compile time.

```c
uint32_t parse(char * p, size_t n) {
    uint32_t ret = 0;
    for (size_t i = 0; (i < n) && p[i]; i++) {
        ret = 10 * ret + (p[i] - '0');
    }
    return ret;
}
```

Figure 7: Parsing in C

```
pub func parser.parse?(src: base.io_reader) {
    var c : base.u8
    while true {
        c = args.src.read_u8?()
        if c == 0 {
            return ok
        }
        if (c < 0x30) or (0x39 < c) { // '0' and '9' in ASCII
            return "#not a digit"
        }
        //Rebase from ASCII to numeric value
        c -= 0x30

        if this.val < 429_496729 {
            this.val = (10 * this.val) + (c as base.u32)
            continue
        } else if (this.val > 429_496729) or (c > 5) {
            return "#too large"
        }

        this.val = (10 * this.val) + (c as base.u32)
    } endwhile
}
```

Figure 8: Parsing in Wuffs

5.2 Wuffs Implementation

In this section, we discuss the mechanics of the Wuffs extension. For more details, examine the concrete syntax file for the `ABLE`C-wuffs extension. The extension is surprisingly simple, given the benefits it adds to a project.

The extension simply looks for a block of code delineated by the markers `WUFFS` and `WUFFS_END`. The code within these two markers is assumed to be valid Wuffs code. This code is loaded into its own `.wuffs` file, which is then compiled using the Wuffs compiler into a C file. In the original `ABLE`C file, we replace the original block of Wuffs code with boilerplate code that includes linking the new C file with the Wuffs functions, as well as calling those functions to parse the input, store the input, and validate the results. Note in the code snippet of the parse function using the extension that all of the Wuffs code is within the larger `parse()` function. This means that we need to do minimal editing of the rest of the file. We can retain the already written functions provided out-of-the-box by the server. In fact, in this case, the only function from the `parse.c` file that needs to be modified is the primary `parse()` function. The end programmer does not need to concern themselves with writing the C code to call any Wuffs functions – the extension is able to take care of inserting those function calls for us, allowing the programmer to simply write the parsing details in Wuffs.

5.3 Wuffs Extension

Below, we have two code snippets of the parsing function from the IRC server. The first snippet, in Figure 9, is from the original server code. It contains a C implementation of the parser. The second code snippet in Figure 10 is the Wuffs implementation of the parser in our extension, with some code removed for brevity.

We consider the differences between these two implementations. Obviously, there is a difference in syntax between the two languages. Beyond that, however, we consider the differences between the two implementations. First, we must primarily note that the C

```
1  GLOBAL bool Parse_Request( CONN_ID Idx, char *Request ){
2          REQUEST req;
3          char *start, *ptr;
4          bool closed;
5
6  //Code omitted for brevity
7
8          if (Request[0] == ':') {
9                  req.prefix = Request + 1;
10                 ptr = strchr( Request, ' ' );
11                 if( ! ptr ) {
12                         return Conn_WriteStr(Idx, "ERROR :Prefix without command");
13                 }
14                 *ptr = '\0';
15                 start = ptr + 1;
16         } else start = Request;
17         ptr = strchr( start, ' ' );
18         if( ptr ) {
19                 *ptr = '\0';
20         }
21         req.command = start;
22         if( ptr ) {
23                 start = ptr + 1;
24                 while( start ) {
25                         if( start[0] == ':' ) {
26                                 req.argv[req.argc] = start + 1;
27                                 ptr = NULL;
28                         } else {
29                                 req.argv[req.argc] = start;
30                                 ptr = strchr( start, ' ' );
31                                 if( ptr ) {
32                                         *ptr = '\0';
33                                 }
34                         }
35                         req.argc++;
36                         if( start[0] == ':' ) break;
37                         if( req.argc > 14 ) break;
38                         if( ptr ) start = ptr + 1;
39                         else start = NULL;
40                 }
41         }
42         if(!Validate_Prefix(Idx, &req, &closed))
43                 return !closed;
44         if(!Validate_Command(Idx, &req, &closed))
45                 return !closed;
46         if(!Validate_Args(Idx, &req, &closed))
47                 return !closed;
48
49         return Handle_Request(Idx, &req);
50 } /* Parse_Request */
```

Figure 9: The parse function before adding the Wuffs extension

```
1  GLOBAL bool Parse_Request( CONN_ID Idx, char *Request ) {
2    WUFFS
3      pub struct parser?(
4        prefix: slice base.u8, cmd : slice base.u8,
5        args : array[15] slice base.u8, argc : base.u32[..=15]
6      )
7      pub func parser.parse?(src: base.io_reader) {
8        var c : base.u8 = args.src.read_u8?() //Code removing whitespace omitted
9        var i : base.u32[..15] = 0
10       var extra : base.u32
11       if c == ':' {
12         extra = 0
13         while true {
14           c = args.src.read_u8?()
15           if c == ' ' {
16             args.src.limited_copy_u32_to_slice!(up_to: extra, s: this.prefix)
17             break
18           }
19           else { extra += 1 }
20         }endwhile
21         c = args.src.read_u8?() //Prep for reading command
22       } else { //Handle command
23         extra = 1
24         while true {
25           c = args.src.read_u8?()
26           if c == ' ' {
27             args.src.limited_copy_u32_to_slice!(up_to: extra, s: this.cmd)
28             break
29           } else { extra += 1 }
30         }endwhile
31       }
32       c = args.src.read_u8?()
33       if c == 0{ return } //No more args, we are finished
34       if c == ' '{ /* Code moving past whitespace omitted */ }
35       else { //At least one arg
36         extra = 1
37         while.loop1 true {
38           while.loop2 true {
39             c = args.src.read_u8?()
40             if c == 0 { break.loop1 } // c is null, we are finished parsing
41             if c == ' ' {
42               args.src.limited_copy_u32_to_slice!(up_to: extra, s: this.args[i])
43               extra = 0
44               i += 1
45               if i >= 14{ break.loop1 }
46               else{ break.loop2 }
47             } else { extra += 1 }
48           }endwhile.loop2
49         }endwhile.loop1
50         this.argc = i+1
51       }
52     } WUFFS_END
53 } // Parse_Request
```

Figure 10: The parse function after adding the Wuffs extension

implementation has no guarantees of safety. If the program compiles, all we can be sure of is that there are no syntax errors in the file. We have no guarantees that, given some malicious input string, we won't be vulnerable to any attacks. In our Wuffs implementation, this code is sent off to the Wuffs compiler during compilation. The Wuffs compiler is able to tell us whether we are vulnerable to buffer overflows or other attacks. This is the largest benefit of using this extension. We get the speed of C code, but the safety and guarantees provided by the Wuffs compiler.

6 Conclusion

We have seen several extensions implemented within the example of an IRC Server written in C. Previously, we have been mostly focused on simply examining the differences between the original and modified servers, without passing judgement on whether those differences are benefits or drawbacks. Within this section, we will focus more on the benefits and drawbacks of the implemented extensions. We finish with a brief discussion of whether these extensions are realistically applicable in a real-world piece of software.

6.1 Non-null Pointers

Consider the extension implementing compile-time and runtime checking for null pointer dereferences, discussed above in section 3. This extension, while relatively simple in its final form, is actually incredibly useful. When compared with a C implementation of a function that takes pointers as arguments, this extension offers immense upside. The programmer no longer needs to worry about checking every pointer for null dereferences at the beginning of the function.

One additional benefit of this extension is the possibility of guaranteeing that the pointer returned from a function is not null. Again, this would allow the programmer to worry much less about checking that pointers are safely dereferenced. This would be particularly useful in contexts where we are not easily able to restart a program if it aborts due to a null pointer dereference; for instance, a singular server running communication between multiple clients (like the IRC server example discussed throughout this paper) is not easily able to restart if it is forced to abort due to an unexpected null pointer dereference. If this were to happen, all of the data stored on the server (messages, channels, and users, as well as their associated histories) would be lost.

There are no readily apparent drawbacks to using our non-null extension, except for the idea that this extension likely will result in minor slowdowns to code. While not tested using

any sort of reliable benchmark, consider the example illustrated in Figures 1, 2, 3, and 4. In particular, we consider Figures 1 and 4. These figures contain the C code that is being compiled and run by the host machine. We observe that Figure 1 contains 6 executable lines of code. Figure 4, on the other hand, contains significantly more lines of code that can be run. This would seem to indicate, at least upon a cursory examination, that the code in Figure 4 would run slightly slower, simply because there are more instructions to execute, though the difference in speed would likely be negligible in this application.

6.2 Asynchronous I/O

Next, we consider the extension implementing nicer facilities for asynchronous I/O than the syntax provided by plain C code. First, we must note that this extension provides only repackaged syntax, not new functionality, when compared with the `epoll` API. The one major benefit provided by this extension, however, is the increase in readability of the code. We no longer must deal directly with the intricacies of the `epoll` API and the concept of *events*. Instead, the extension allows us to focus on the conceptual ideas in the code; that is, we simply want some function task to be run asynchronously, and we don't need to worry about modifying which tasks the `epoll` instance is tracking manually.

One drawback to our asynchronous extension is that we potentially lose some of the finer grains of control provided to us by the raw `epoll` API. As an example, consider Figure 5, specifically line 45. In this code pulled from the original IRC server, we see a call to `io_event_add` with the constant `IO_WANTREAD`. This indicates that we are starting an asynchronous task that specifically wants to perform `read` operations in I/O. Consider the version of this code with the extension implemented in Figure 6, again at line 45. Here, we no longer have the ease of simply specifying that we want to perform an asynchronous `read` operation. Instead, we must define the specific function we want to run.

While this lack of control could potentially be improved using a more complex extension, we did not consider such an extension within this paper. Additionally, adding complexity

to the extension is perhaps antithetical to the idea of extensions in the first place. We wish to abstract away certain aspects of the host language using new syntax, so while we can achieve a more full control using more complexity, this added complexity may become more confusing than simply using the original host language to achieve the same effect.

6.3 Wuffs

Finally, we consider the Wuffs extension that allows for more secure parsing. This extension is a wonderful example of the benefits of using extensible programming languages. We introduce a secure way of parsing using the Wuffs language, leveraging this new tool to write a secure parsing function. The benefits to using this extension are vast, especially when compared with the original C parsing function. The guarantees offered by Wuffs allow us to be certain that our program has no vulnerabilities to a variety of attacks. Additionally, with this extension, we no longer must worry about manually writing out the code to make the necessary calls to the Wuffs code after it has been compiled into C. Instead, the extension is able to take care of inserting these function calls as necessary. Another benefit of this extension is that we do not need to worry about integrating a separate Wuffs file. We are instead able to write the Wuffs code *in the same file* as the C code that will interact with the Wuffs code.

Even with the added guarantees provided by the Wuffs compiler (requiring more checks than would be required in C code), the Wuffs extension will not have a noticeable slowdown, and could even offer a moderate speed-up in certain cases. For more details, see the benchmark documentation in the Wuffs repository for a comparison of various Wuffs libraries and default C implementations of those same libraries [**5**].

6.4 Applications in Other Software

We have seen the benefits brought by the extensions introduced in this paper. But are extensions a realistic way of writing software for any real-world application? The work

within this paper suggests that extensions do indeed provide concrete benefits in real-world applications. We have seen that extensions, particularly those introducing compile-time checking of various language features, are incredibly useful in saving programmers time and energy when writing new software. It is not unreasonable to assume that an extension similar to the non-null extension discussed in this paper would be incredibly useful in a large-scale C project.

Additionally, extensions similar to the Wuffs extension would be incredibly useful in scenarios where we are dealing with safety-critical pieces of code. Any code that deals with possibly malicious inputs would benefit from the guarantees provided by Wuffs. From buffer overflows to integer overflows or underflows, the Wuffs extension guarantees that we are protected from these attacks, allowing programmers to no longer worry about writing more complicated safety checks in C.

7 Future Work

There are several possible ways of improving these extensions in the future. First, we consider the extension providing improved asynchronous I/O facilities. One possible way of improving this extension would be to expand the functionality beyond just a simple translation of `spawn` to `io_event_add` and `await` to `io_dispatch`. While this functionality does provide better syntax for this project, having improved syntax to allow for more complicated constructs would be useful. One potential direction for this extension to go would be to follow a similar pattern to the `select` construct in Go. This allows for programmers to wait for multiple communication operations to finish and then execute a set of instructions based on which communication finishes first. This, implemented as an extension, would allow programmers to deal with conditions where we want to perform a read or write as soon as possible, as opposed to simply doing so asynchronously.

We now consider what further improvements could be made to our Wuffs extension. First, we note that this the parser we have implemented in this project is relatively simple, only parsing text as input, where other parsers must parse much more complicated entities like images, audio files, and other file formats. In a more expansive project, it would be a nice feature for a programmer to be able to include in multiple locations different Wuffs blocks. To do this, we would need to allow the programmer to specify the Wuffs package name, as well as what calls to Wuffs functions the programmer would like to make. This would potentially eliminate some of the automation from the Wuffs extension, but it would still provide the same advantages (just to a lesser extent) over placing the Wuffs code in a separate file that must be integrated into the project separately.

References

[1] A. Barton, "Next generation IRC daemon2," ngIRCd, 2001. [Online]. Available: https://ngircd.barton.de/. [Accessed: 22-Nov-2022].

[2] Councilman, Aaron. (2021). An Extensible Implementation-Agnostic Parallel Programming Framework for C in ableC. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/220246.

[3] E. Van Wyk, D. Bodin, J. Gao, and L. Krishnan, "Silver: An extensible attribute grammar system," Electronic Notes in Theoretical Computer Science, vol. 203, no. 2, pp. 103–116, Jan. 2010.

[4] E. Van Wyk, O. de Moor, K. Backhouse, and P. Kwiatkowski "Forwarding in attribute grammars for modular language design," Lecture Notes in Computer Science, pp. 128–142, Apr. 2002.

[5] Google, "Google/WUFFS: Wrangling untrusted file formats safely," GitHub. [Online]. Available: https://github.com/google/wuffs. [Accessed: 18-Nov-2022].

[6] M. D. McIlroy, "Macro instruction extensions of compiler languages," Communications of the ACM, vol. 3, no. 4, pp. 214–220, 1960.

[7] Oikarinen, J. and D. Reed, "Internet Relay Chat Protocol", RFC 1459, DOI 10.17487/RFC1459, May 1993, ¡https://www.rfc-editor.org/info/rfc1459¿.

[8] T. Kaminski, L. Kramer, T. Carlson, and E. Van Wyk, "Reliable and automatic composition of language extensions to C: The ABLEC extensible language framework," Proceedings of the ACM on Programming Languages, vol. 1, no. OOPSLA, pp. 1–29, 2017.