BLASTing Linux Code*

Jan Tobias Mühlberg and Gerald Lüttgen

Department of Computer Science, University of York, York YO10 5DD, U.K.
<muehlber|luettgen>@cs.york.ac.uk

Abstract. Computer programs can only run reliably if the underlying operating system is free of errors. In this paper we evaluate, from a practitioner's point of view, the utility of the popular software model checker BLAST for revealing errors in Linux kernel code. The emphasis is on important errors related to memory safety in and locking behaviour of device drivers. Our conducted case studies show that, while BLAST's abstraction and refinement techniques are efficient and powerful, the tool has deficiencies regarding usability and support for analysing pointers, which are likely to prevent kernel developers from using it.

1 Introduction

Today's application software critically depends on the reliability, safety and security of the underlying operating system (OS). However, due to their complicated task of managing a system's physical resources, OSs are difficult to develop and even more difficult to debug. Quite frequently major errors stay undiscovered until they are exploited in security attacks or are found "by accident".

In recent years, automatic approaches to discover OS bugs via runtime checks or source code analysis have been explored. Despite the fact that many of these approaches do not focus on an exhaustive analysis, they still helped developers to detect hundreds of safety problems in the Linux and BSD OS kernels. Most of the programming errors found were either related to *memory safety* or incorrect *locking behaviour* [6]. Here, "memory safety" typically is interpreted as the property that an OS component never de-references an invalid pointer, since this would cause the program to end up in an undefined state. "Correct locking behaviour" means that functions that ensure mutual exclusion on the physical resources of a system are called in a way that is free of deadlocks and starvation. Both classes of problems are traceable by checking whether an OS component complies with basic usage rules of the program interface provided by the kernel.

Software model checking. By having the potential of being exhaustive and fully automatic, *model checking*, in combination with *abstraction* and *refinement*, is a successful technique used in software verification [7]. Intensive research in this area has resulted in software model checkers like Bandera [9] for Java programs or SLAM/SDV [1], MAGIC [5] and BLAST [16] (*Berkeley Lazy Abstraction*

* Research funding was provided by the EPSRC under grant GR/S86211/01.

Software verification Tool) for analysing C source code. The major advantage of these tools over model-based model checkers such as Spin [17] is their ability to automatically abstract a model from the source code of a given program. User interaction should then only be necessary in order to provide the model checker with a specification, against which the program can be checked. Since complete formal specifications are not available for most programs, verification will usually be relative to a partial specification that covers the usage rules of the *Application Program Interface* (API) used by the program. However, up to now all releases of SLAM are restricted to verifying properties for Microsoft Windows device drivers and do not cover memory safety problems [19], while BLAST and MAGIC are able to verify a program against a user defined temporal safety specification and thus allows checking of arbitrary C source code.

The* BLAST *toolkit . This popular toolkit implements an advanced abstraction algorithm, called "lazy abstraction" [15], for building a model of some C source code, and model-checking algorithm for checking whether some specified label placed in the source code is reachable. This label can either be automatically introduced by instrumenting the source with an explicit temporal safety specification, be added via `assert()` statements, or be manually introduced into the source. In any case, the input source file needs to be preprocessed using a standard C preprocessor like `gcc`. In this step, all header and source files included by the input file under consideration are merged into one file. It is this preprocessed source code that is passed to BLAST to construct and verify a model using *predicate abstraction*.

This paper. In this paper we investigate to which extent software model checking as implemented in BLAST can aid a practitioner during OS software development. To do so, we analyse whether BLAST is able to detect errors that have been reported for recent releases of the Linux kernel. We consider programming errors related to *memory safety* (cf. Sec. 3) and *locking behaviour* (cf. Sec. 4). The code examples utilised in this paper are taken from releases 2.6.13 and 2.6.14 of the Linux kernel. They have been carefully chosen by searching the kernel's change log for fixed memory problems and fixed deadlock conditions, in a way that the underlying problems are representative for memory safety and locking behaviour as well as easily explainable without referring to long source code listings.[1] Our studies use version 2.0 of BLAST, which was released in October 2005.

The focus of our work is on showing at what scale a give problem statement and a program's source code need to be adapted in order to detect an error. We discuss how much work is required to find a certain usage rule violation in a given snippet of a Linux driver, and how difficult this work is to perform in BLAST. Due to space constraints, we cannot present all of our case studies in full here; however, all files necessary to reproduce our results can be downloaded from `www.cs.york.ac.uk/~muehlber/blast/`.

[1] All source code used is either included or referenced by a *commit key* as provided by the source code management system *git* which is used in the Linux kernel community; see `www.kernel.org` for further information on *git* and Linux.

***Related studies with* BLAST.** BLAST has been applied for the verification of memory safety as well as locking properties before [3,13,16,14]. In [3], the use of CCURED [21] in combination with BLAST for verifying memory safety of C source code is explained. This is done by inserting additional runtime checks at all places in the code where pointers are de-referenced. BLAST is then employed to check whether the introduced code is reachable or can be removed again. The approach focuses on ensuring that only valid pointers are de-referenced along the execution of a program, which is taken to mean that pointers must not equal `NULL` at any point at which they are de-referenced. However, invalid pointers in C do not necessarily equal `NULL` in practise. In contrast to [3], we will interpret pointer invalidity in a more general way and conduct our studies on real-world examples rather than constructed examples.

A methodology for verifying and certifying systems code on a simple locking problem is explained in [16], which deals with the spinlock interface provided by the Linux kernel. *Spinlocks* ensure that a kernel process can spin on a CPU without being preempted by another process. The framework studied in [16] is used to prove that calls of `spin_lock()` and `spin_unlock()` in Linux device drivers always alternate. In contrast to this work, our case studies will be more detailed and thereby will be providing further insights into the usability of BLAST.

2 Programming Errors in OS Code

There is quite a long list of commonly found OS errors. While most of them mainly affect a system's safety, others have a security-related background. An insightful study of OS errors has been published in [6]; see Table 1 for a summary of its results. The study shows that the majority of programming errors in OS code can be found in device drivers. Its authors highlight that most errors are related to problems causing either deadlock conditions or driving the system into undefined states by de-referencing invalid pointers.

Although memory safety problems have a direct impact on an OS's reliability, API rules for OS kernels are usually described in an informal way. For example, in the Linux device driver handbook [8, p. 61] it is stated that one "should never pass anything to *kfree* that was not obtained from *kmalloc*" since, otherwise, the system may behave in an undefined way. The functions `kmalloc()` and `kfree()` are kernel-space functions which are used to dynamically allocate and de-allocate memory, respectively. Another common example are buffer overrun errors, where data is written beyond the size of an allocated area of memory, thus overwriting unrelated data.

Correct locking of resources is another major issue causing problems in OS code. As shown in [6], deficiencies resulting in deadlocks in the Linux and BSD kernels make up a large amount of the overall number of errors found. In the documentation explaining the API of the Linux kernel, quite strict rules about the proper use of functions to lock various resources are stated. For example, in [8, p. 121], one of the most basic rules is given as follows: "Neither semaphores nor spinlocks allow a lock holder to acquire the lock a second time; should

Table 1. Results of an empirical study of OS errors [6]

% of Bugs	Rule checked
63.1%	**Bugs related to memory safety**
38.1%	Check potentially NULL pointers returned from routines.
9.9%	Do not allocate large stack variables (> 1K) on the fixed-size kernel stack.
6.7%	Do not make inconsistent assumptions about whether a pointer is NULL.
5.3%	Always check bounds of array indices and loop bounds derived from user data.
1.7%	Do not use freed memory.
1.1%	Do not leak memory by updating pointers with potentially NULL realloc return values.
0.3%	Allocate enough memory to hold the type for which you are allocating.
33.7%	**Bugs related to locking behaviour**
28.6%	To avoid deadlock, do not call blocking functions with interrupts disabled or a spinlock held.
2.6%	Restore disabled interrupts.
2.5%	Release acquired locks; do not double-acquire locks.
3.1%	**Miscellaneous bugs**
2.4%	Do not use floating point in the kernel.
0.7%	Do not de-reference user pointers.

you attempt to do so, things simply hang." The rational for this lies in the functionality provided by spinlocks: a kernel thread holding a lock is spinning on one CPU and cannot be preempted until the lock is released. Another important rule is that any code holding a spinlock cannot relinquish the processor for anything except for serving interrupts; especially, the thread must never sleep because the lock might never be released in this case [8, p. 118].

3 Checking Memory Safety

This section focuses on using Blast for checking usage rules related to memory safety, for which we have analysed several errors in different device drivers. The examples studied by us include use-after-free errors in the kernel's SCSI[2] and InfiniBand[3] subsystems. The former is the *small computer system interface* standard for attaching peripheral devices to computers, while the latter is an industry standard designed to connect processor nodes and I/O nodes to form a system area network. In each of these examples, an invalid pointer that is not `NULL` is de-referenced, which causes the system to behave in an undefined way. This type of bug is not covered by the work on memory safety of Beyer et al. in [3] and cannot easily be detected by runtime checks.

[2] Commit `2d6eac6c4fdaa69656d66c80754d267be233cc3f`.
[3] Commit `d0743a5b7b837334cb414b773529d51de3de0471`.

The example we will study here in detail is a use-after-free error spotted by the Coverity source code analyser (`www.coverity.com`) in the I2O subsystem of the Linux kernel (cf. Sec. 3.1). To check for this bug in Blast we first specify a temporal safety specification in the Blast specification language. Taking this specification, Blast is supposed to automatically generate an instrumented version of the C source code for analysis (cf. Sec. 3.2). However, due to an apparent bug in Blast, this step fails for our example, and we are therefore forced to manually instrument our code by inserting `ERROR` labels at appropriate positions (cf. Sec. 3.3). However, it will turn out that Blast does not track important operations on pointers, which is not mentioned in Blast's user manual and without which our example cannot be checked (cf. Sec. 3.4).

3.1 The I2O Use-After-Free Error

The I2O subsystem bug of interest to us resided in lines 423–425 of the source code file `drivers/message/i2o/pci.c`. The listing in Fig. 1 is an abbreviated version of the file `pci.c` before the bug was fixed. One can see that function `i2o_iop_alloc()` is called at line 330 of the code extract. This function is defined in `drivers/message/i2o/iop.c` and basically allocates memory for an `i2o_controller` structure using `kmalloc()`. At the end of the listing, this memory is freed by `i2o_iop_free(c)`. The bug in this piece of code arises from the call of `put_device()` in line 425, since its parameter `c->device.parent` causes an already freed pointer to be de-referenced. The bug has been fixed in commit `d2b0e84d195a341c1cc5b45ec2098ee23bc1fe9d`, by simply swapping lines 424 and 425 in the source file.

```
drivers/message/i2o/pci.c:
 300  static int __devinit
       i2o_pci_probe(
       struct pci_dev *pdev,
 301   const struct pci_device_id
       *id)
 302  {
 303   struct i2o_controller *c;
```

```
330   c = i2o_iop_alloc();

423   free_controller:
424    i2o_iop_free(c);
425    put_device(
         c->device.parent);

432  }
```

Fig. 1. Extract of `drivers/message/i2o/pci.c`.

This bug offers various different ways to utilise Blast. A generic temporal safety property for identifying bugs like this would state that *any pointer that has been an argument to* ***`kfree()`*** *is never used again* unless it has been reallocated. A probably easier way would be to check whether *the pointer* `c` *in* ***`i2o_pci_probe()`*** *is never used again after* ***`i2o_iop_free()`*** *has been called* with `c` as its argument. Checking the first, more generic property would require us to put function definitions from other source files into `pci.c`, since Blast considers only functions that are available in its input file. Therefore, we focus on verifying the latter property.

Checking for violations even of the latter, more restricted property will lead to a serious problem. A close look at the struct `i2o_controller` and its initialisation in the function `i2o_iop_alloc()` reveals that `i2o_controller` contains a function pointer which can be used as a "destructor". As is explained in BLAST's user manual, the "current release does not support function pointers"; they are ignored completely. Further, the manual states that "correctness of the analysis is then modulo the assumption that function pointer calls are irrelevant to the property being checked." This assumption is however not always satisfied in practise, as we will see later in our example.

3.2 Verification With a Temporal Safety Specification

Ignoring the function pointer limitation, we developed the temporal safety specification presented in Fig. 2. The specification language used by BLAST is easy to understand and allows the assignment of status variables and events. In our specification we use a global status variable `allocstatus_c` to cover the possible states of the struct `c` of our example, which can be set to 0 meaning "not allocated" and 1 meaning "allocated". Furthermore, we define three events, one for each of the functions `i2o_iop_alloc()`, `i2o_iop_free()` and `put_device()`. All functions have special preconditions and calling them may modify the status of `c`. The special token `$?` matches anything. Intuitively, the specification given in Fig. 2 states that `i2o_iop_alloc()` and `i2o_iop_free()` must be called alternately, and `put_device()` must only be called when `c` has not yet been freed. Note that this temporal safety specification does not cover the usage rule for `i2o_iop_free()` and `put_device()` in general. We are using one status variable to guard calls of `i2o_iop_free()` and `put_device()` regardless of its arguments. Hence, the specification will work only as long as there is only one pointer to an `i2o_controller` structure involved.

```
global int allocstatus_c = 0;

event
 {
  pattern { $? = i2o_iop_alloc(); }
  guard   { allocstatus_c == 0 }
  action  { allocstatus_c = 1; }
 }
```

```
event
 {
  pattern { i2o_iop_free($?); }
  guard   { allocstatus_c == 1 }
  action  { allocstatus_c = 0; }
 }

event
 {
  pattern { put_device($?); }
  guard   { allocstatus_c == 1 }
 }
```

Fig. 2. A temporal safety specification for `pci.c`.

Using the specification of Fig. 2, BLAST should instrument a given C input file by adding a global status variable and error labels for all violations of the

preconditions. The instrumentation is done by the program `spec.opt` which is part of the BLAST distribution. For our example taken from the Linux kernel, we first obtained the command used by the kernel's build system to compile `pci.c` with `gcc`. We appended the option `-E` to force the compilation to stop after preprocessing, resulting in a C source file containing all required parts of the kernel headers. This step is necessary since BLAST cannot know of all the additional definitions and include paths used to compile the file. Unfortunately, it expands `pci.c` from 484 lines of code to approximately 16k lines, making it difficult to find syntactical problems which BLAST cannot deal with. Despite spending a lot of effort in trying to use `spec.opt`, we never managed to get this work. The program mostly failed with unspecific errors such as `Fatal error: exception Failure("Function declaration not found")`. Finding such an error in a huge source without having a line number or other hint is almost impossible, especially since `gcc` compiles the file without any warning. We constructed several simplifications of the preprocessed file in order to trace the limitations of `spec.opt`, but did not get a clear indication of what the source is. We suspect it might be a problem with parsing complex data structures and inline assembly imported from the Linux headers.

Given the bug in BLAST and in order to demonstrate that our specification indeed covers the programming error in `pci.c`, we developed a rather abstract version of `pci.c` which is shown in Fig. 3. Using this version and the specification of Fig. 2, we were able to obtain an instrumented version of our source code without encountering the bug in `spec.opt`. Running BLAST on the instrumented version then produced the following output:

```
$ spec.opt test2.spc test2.c
[...]
$ pblast.opt instrumented.c
[...]
Error found! The system is unsafe :-(
```

In summary, the example studied here shows that the specification used in this section is sufficient to find the bug. However, the approach required by BLAST has several disadvantages. Firstly, it is not automatic at all. Although we ended up with only a few lines of code, it took quite a lot of time to produce this code by hand and to figure out what parts of the original `pci.c` are accepted by BLAST. Secondly, the methodology only works if the bug is known beforehand, hence we did not learn anything new about unwanted behaviour of this driver's code. We needed to simplify the code to an extent where the relation to the original source code may be considered as questionable. The third problem lies in the specification used. Since it treats the allocation and de-allocation as something similar to a locking problem, we would not be able to use it in a piece of code that refers to more than one dynamically allocated object. A more generic specification must be able to deal with multiple pointers. According to [2], such a generic specification should be possible to write by applying a few minor modifications such as defining a "shadow" control state and replacing `$?`

```
test2.h:
#include <stdio.h>
#include <stdlib.h>

typedef struct device
{
 int parent;
} device;

typedef struct i2o_controller
{
 struct device device;
} i2o_controller;

i2o_controller *i2o_iop_alloc
 (void);
void i2o_iop_free
 (i2o_controller *c);
void put_device (int i);
```

```
test2.c:
#include "test2.h"

i2o_controller *i2o_iop_alloc
 (void)
{ i2o_controller *c;
  c = malloc(
   sizeof(struct i2o_controller));
  return (c); }

void i2o_iop_free
 (i2o_controller *c)
{ free (c); }

void put_device (int i) { }

int main (void)
{ i2o_controller *c;
  c = i2o_iop_alloc ();
  i2o_iop_free (c);
  put_device (c->device.parent);
  return (0); }
```

Fig. 3. Manual simplification of `pci.c`.

with `$1`. However, in practise the program generating the instrumented C source file failed with obscure error messages.

3.3 Verification Without a Temporal Safety Specification

Since Blast could not deal with verifying the original `pci.c` using an explicit specification of the use-after-free property, we will now try and manually instrument the source file so that our bug can be detected whenever an `ERROR` label is reachable.

When conducting our instrumentation, the following modifications were applied by hand to `pci.c` and related files:

1. A variable `unsigned int alloc_status` was added to the definition of `struct i2o_controller` in `include/linux/i2o.h`.
2. The prototypes of `i2o_iop_alloc()` and `i2o_iop_free()` were removed from `drivers/message/i2o/core.h`.
3. The prototype of `put_device()` was deleted from `include/linux/device.h`.
4. C source code for the functions `put_device()`, `i2o_iop_free()`, `i2o_iop_release()` and `i2o_iop_alloc()` was copied from `iop.c` and `drivers/base/core.c` into `pci.c`. The functions were modified such that the new field `alloc_status` of a freshly allocated `struct i2o_controller` is set to 1 by `i2o_iop_alloc()`. `i2o_iop_free()` no longer de-allocates the structure but checks whether `alloc_status` equals 1 and sets it to 0; otherwise, it jumps

to the `ERROR` label. `put_device()` was modified to operate on the whole `struct i2o_controller` and jumps to `ERROR` if `alloc_status` equals 0.

By feeding these changes into the model checker it is possible to detect duplicate calls of `i2o_iop_free()` on a pointer to a `struct i2o_controller`, as well as calls of `put_device()` on a pointer that has already been freed. Even calls of `i2o_iop_free()` and `put_device()` on a pointer that has not been allocated with `i2o_iop_alloc()`, should result in an error report since nothing can be said about the status of `alloc_status` in such a case.

After preprocessing the modified source files and running BLAST, we get the output `"Error found! The system is unsafe :-("`. Even after we reduced the content of `i2o_pci_probe()` to something quite similar to the `main()` function shown in Fig. 3 and after putting the erroneous calls of `put_device()` and `i2o_iop_free()` in the right order, the system was still unsafe from BLAST's point of view. It took us some time to figure out that BLAST does not appear to consider the content of pointers at all.

3.4 The Problem with BLAST and Pointers

We demonstrate this apparent shortcoming of BLAST regarding handling pointers by means of another simple example, for which BLAST fails in tracing values behind pointers over function calls.

```
test5.c:
 1  #include <stdlib.h>
 2
 3  typedef struct example_struct
 4   {
 5    void   *data;
 6    size_t size;
 7   } example_struct;
 8
 9
10  void init (example_struct *p)
11   {
12    p->data = NULL;
13    p->size = 0;
14
15    return;
16   }
17
18  int main (void)
19   {
20    example_struct p1;
21
22    init (&p1);
23    if (p1.data != NULL ||
          p1.size != 0)
24     { goto ERROR; }
25     else
26     { goto END; };
27
28  ERROR:
29    return (1);
30
31  END:
32    return (0);
33   }
```

Fig. 4. An example for pointer passing.

As can be seen in the code listing of Fig 4, label `ERROR` can never be reached in this program since the values of the components of our struct are explicitly set by function `init()`. However, BLAST produces the following output:

```
$ gcc -E -o test5.i test5.c
$ pblast.opt test5.i
[...]
Error found! The system is unsafe :-(
Error trace:
23 :: 23: Pred((p1@main).data!=0) :: 29
-1 :: -1: Skip :: 23
10 :: 10: Block(Return(0);) :: -1
12 :: 12: Block(* (p@init ).data = 0;* (p@init ).size = 0;) :: 10
22 :: 22: FunctionCall(init(&(p1@main))) :: -1
-1 :: -1: Skip :: 22
 0 ::  0: Block(Return(0);) :: -1
 0 ::  0: FunctionCall (__BLAST_initialize_test5.i()) :: -1
```

This counterexample shows that Blast does not correlate the pointer `p` used in `init()` and the struct `p1` used in `main()`, and assumes that the `if` statement in line 23 evaluates to true. After adding a line "`p1.data = NULL; p1.size = 0;`" before the call of `init()`, Blast claims the system to be safe, even if we modify `init()` to reset the values so that they differ from `NULL` (and `0`).

We were able to reproduce this behaviour in similar examples with pointers to integer values and arrays. Switching on the BDD-based alias analysis implemented in Blast also did not solve the problem. The example shows that Blast does not only ignore function pointer calls as stated in its user manual, but appears to assume that all pointer operations have no effect. This limitation is not documented in the Blast manual and renders Blast almost unusable for the verification of properties related to our understanding of memory safety.

3.5 Results

Our experiments on memory safety show that Blast is able to find the programming error discovered by the Coverity checker. Out of eight examples, we were able to detect two problems after minor modifications to the source code, and three after applying manual abstraction. Three further programming errors could not be traced by using Blast. Indeed, Blast has some major restrictions. The main problem is that Blast ignores variables addressed by a pointer. As stated in its user manual, Blast assumes that only variables of the same type are aliased. Since this is the case in our examples, we initially assumed that our examples could be verified with Blast, which is not the case. Moreover, we encountered bugs and deficiencies in `spec.opt` which forced us to apply substantial and time consuming modifications to source code. Most of these modifications and simplifications would require a developer to know about the error in advance. Thus, from a practitioner's point of view, Blast is not of much help in finding unknown errors related to memory safety. However, it needs to be mentioned that Blast was designed for verifying API usage rules of a different type than those required for memory safety. More precisely, Blast is intended for proving the adherence of pre- and post-conditions denoted by integer values and for ensuring API usage rules concerning the order in which certain functions are called, regardless of pointer arguments, return values and the effects of aliasing.

4 Checking Locking Properties

Verifying correct locking behaviour is something used in almost all examples provided by the developers of BLAST [2,16]. In [16], the authors checked parts of the Linux kernel for correct locking behaviour while using the *spinlock* API and stated that BLAST showed a decent level of performance during these tests. Spinlocks provide a very simple but quite efficient locking mechanism to ensure, e.g., that a kernel thread may not be preempted while serving interrupts. The kernel thread acquires a certain lock by calling `spin_lock(l)`, where `l` is a previously initialised pointer to a struct `spinlock_t` identifying the lock. A lock is released by calling `spin_unlock()` with the same parameter. The kernel provides a few additional functions that control the interrupt behaviour while the lock is held. By their nature, spinlocks are intended for use on multiprocessor systems where each resource may be associated with a special spinlock, and where several kernel threads need to operate independently on these resources. However, as far as concurrency is concerned, uniprocessor systems running a preemptive kernel behave like multiprocessor systems.

```
global int lockstatus = 2;

event
 {
  pattern { spin_lock_init($?); }
  guard   { lockstatus == 2 }
  action  { lockstatus = 0; }
 }

event
 {
  pattern { spin_lock($?); }
  guard   { lockstatus == 0 }
  action  { lockstatus = 1; }
 }
```

```
event
 {
  pattern { spin_unlock($?); }
  guard   { lockstatus == 1 }
  action  { lockstatus = 0; }
 }

event
 {
  pattern { $? = sleep($?); }
  guard   { lockstatus == 0 }
 }
```

Fig. 5. A temporal safety specification for spinlocks.

Finding examples for the use of spinlocks is not difficult since they are widely deployed. While experimenting with BLAST and the spinlock functions on several small components of the Linux kernel we experienced that it performs well with functions using only one lock. We focused on functions taken from the USB subsystem in *drivers/usb/core*. Due to further unspecific parse errors with the program `spec.opt` we could not use a temporal safety specification directly on the kernel source. However, in this case we were able to generate the instrumented source file and to verify properties by separating the functions under consideration from the remaining driver source and by providing simplified header files.

In Fig. 5 we provide our basic temporal safety specification for verifying locking behaviour. Variable `lockstatus` encodes the possible states of a spinlock; the initial value 2 represents the state in which the lock has not been initialised, while 1 and 0 denote that the lock is held or has been released, respectively. The pattern within the specification varies for the different spinlock functions used within the driver source under consideration, and the specification can easily be extended to cover forbidden functions that may sleep. An example for a function `sleep()` is provided in the specification of Fig. 5.

Difficulties arise with functions that acquire more than one lock. Since all spinlock functions use a pointer to a struct `spinlock_t` in order to identify a certain lock, and since the values behind pointers are not sufficiently tracked in BLAST, we were forced to rewrite parts of the driver's source and the kernel's spinlock interface. Instead of the pointers to `spinlock_t` structs we utilise global integer variables representing the state of a certain lock. We have used this methodology to verify an example of a recently fixed deadlock[4] in the Linux kernel's SCSI subsystem. In Fig. 6 we provide an extract of one of the functions modified in the fix. We see that the spinlocks in this example are integrated in more complex data structures referenced via pointers. Even worse, this function calls a function pointer passed in the argument `done` in line 1581, which was the source of the deadlock before the bug was fixed. To verify this special case, removing the function pointer and providing a dummy function `done()` with a precondition assuring that the lock on `shost->host_lock` is not held is needed. However, we were able to verify both the deadlock condition before the fix had been applied, as well as deadlock freedom for the fixed version of the source.

```
1564 int ata_scsi_queuecmd(struct
      scsi_cmnd *cmd, void
      (*done)(struct scsi_cmnd *))
1565 {
1566 struct ata_port *ap;
1567 struct ata_device *dev;
1568 struct scsi_device
      *scsidev = cmd->device;
1569 struct Scsi_Host
      *shost = scsidev->host;
1571 ap = (struct ata_port *)
      &shost->hostdata[0];
1573 spin_unlock(shost->host_lock);
1574 spin_lock(&ap->host_set->lock);

1581 done(cmd);

1597 spin_unlock(&ap->host_set->lock);
1598 spin_lock(shost->host_lock);
1600 }
```

Fig. 6. Extract of `drivers/scsi/libata-scsi.c`.

During our experiments we analysed several other examples of deadlock conditions. The more interesting examples are the spinlock problem explained above, and another one in the SCSI subsystem,[5] as well as a bug in a IEEE1394 driver[6]. We were able to detect the locking problems in all of these examples and proved the fixed source files to be free of these bugs.

[4] Commit `d7283d61302798c0c57118e53d7732bec94f8d42`.
[5] Commit `fe2e17a405a58ec8a7138fee4ebe101858b636e0`.
[6] Commit `910573c7c4aced8fd5f45c334cc67862e3424d92`.

Results. Out of eight examples for locking problems we were able to detect only five. However, when comparing our results with the conclusions of the previous section, BLAST worked much better for the locking properties because it required fewer modifications to the source code. From a practitioner's point of view, BLAST performed acceptable as long as only one lock was involved. After considerable efforts in simplifying the spinlock API — mainly removing the use of pointers and manually adding error labels to the spinlock functions — we also managed to deal with multiple locks. However, we consider it as fairly difficult to preserve the behaviour of functions that may sleep and therefore must not be called under a spinlock. Even for large portions of source code, BLAST returned its results within a few seconds or minutes, on a PC equipped with an AMD Athlon 64 processor running at 2200 MHz and 1 GB of RAM. Hence, BLAST's internal slicing and abstraction techniques work very well.

We have to point out that the code listing in Fig. 6 represents one of the easily understandable programming errors. Many problems in kernel source code are more subtle. For example, calling functions that may sleep is something that needs to be avoided. However, if a driver calls a function not available in source code in the same file as the driver under consideration, BLAST will only be able to detect the problem if there is an event explicitly defined for this function.

5 Issues with BLAST

This section highlights various shortcomings of the BLAST toolkit which we experienced during our studies. We also present ideas on how BLAST could be improved in order to be more useful for OS software verification.

Lack of documentation. Many problems while experimenting with BLAST were caused by the lack of consistent documentation. For example, a significant amount of time could have been saved in our experiments with memory safety, if the BLAST manual would state that almost all pointer operations are ignored. An in-depth discussion of the features and limitations of the alias analysis implemented in BLAST would also be very helpful to have.

Non-support of pointers. The fact that BLAST does not properly support the use of pointers, in the sense of Sec. 3.4, must be considered as a major restriction, and made our experiments with the spinlock API rather difficult. The restriction forces one to carry out substantial and time consuming modifications to source code. Furthermore, it raises the question whether all important predicates of a given program can be preserved in a manual step of simplification. In some of our experiments we simply replaced the pointers used by the spinlock functions with integers representing the state of the lock. This is obviously a pragmatic approach which does not reflect all possible behaviour of pointer programs. However, it turned out that it is expressive enough to cover the usage rules of the spinlock API. As such modifications could be introduced into the source code automatically, we consider them as an interesting extension for BLAST.

The missing support of function pointers has already been mentioned in Sec. 3. It is true that function pointers are often used in both application space and OS development. In most cases their effect on the program execution can only be determined at run-time, not statically at compile-time. Therefore, we assume that simply skipping all calls of function pointers is acceptable for now.

Usability. There are several issues regarding BLAST's usability which are probably easy to fix, but right now they complicate the work with this tool. Basically, if a piece of C source is accepted by an ANSI C compiler, it should be accepted by BLAST rather than raising uninformative error messages.

A nice improvement would be to provide wrapper scripts that automate preprocessing and verification in a way that BLAST can be used with the same arguments as the compiler. It could be even more useful if functions that are of interest but from other parts of a given source tree, would be copied in automatically. Since we obviously do not want to analyse the whole kernel source in a single file, this should be integrated into BLAST's abstraction/model checking/refinement loop.

6 Related Work

Much work on techniques and tools for automatically finding bugs in software systems has been published in recent years.

Runtime analysis. A popular runtime analysis tool which targets memory safety problems is Purify (`www-306.ibm.com/software/awdtools/purify/`). It mainly focuses on detecting and preventing memory corruption and memory leakage. However, Purify and other such tools, including Electric Fence (`perens.com/FreeSoftware/ElectricFence/`) and Valgrind (`valgrind.org`), are meant for testing purposes and thereby only cover the set of program runs specified by the underlying test cases. An exhaustive search of a programs state space, as is done in model checking, is out of the scope of these tools.

Static analysis and abstract interpretation. Static analysis is another powerful technique for inspecting source code for bugs. Indeed, most of the memory safety problems within the examples of this paper had been detected earlier via an approach based on system-specific compiler extensions, known as *meta-level compilation* [11]. This approach is implemented in the tool Coverity (`www.coverity.com`) and was used in [6]. A further recent attempt to find bugs in OS code is based on abstract interpretation [10] and presented in [4]. The authors checked about 700k lines of code taken from recent versions of the Linux kernel for correct locking behaviour. The paper focuses on the kernel's spinlock interface and problems related to sleep under a spinlock. Several new bugs in the Linux kernel were found during the experiments. However, the authors suggest that their approach could be improved by adopting model checking techniques. An overview of the advantages and disadvantages of static analysis versus model checking can be found in [12].

***Case studies with* Blast.** We have already referred to some such case studies in the introduction. Two project reports of graduate students give further details on Blast's practical use. In [20], Mong applies Blast to a doubly linked list implementation with dynamic allocation of its elements and verifies correct allocation and de-allocation. The paper explains that Blast was not powerful enough to keep track of the state of the list, i.e., the number of its elements. Jie and Shivkumar report in [18] on their experience in applying Blast to a user level implementation of a virtual file system. They focus on verifying correct locking behaviour for data structures of the implementation and were able to successfully verify several test cases and to find one new error. However, in the majority of test cases Blast failed due to documented limitations, e.g., by not being able to deal with function pointers, or terminated with obscure error messages. Both studies were conducted in 2004 and thus based on version 1.0 of Blast. As shown in this paper, Blast's current version has similar limitations.

7 Conclusions and Future Work

We exposed Blast to analysing 16 different OS code examples of programming errors related to memory safety and locking behaviour. Details of the examples which we could not show here due to a lack of space, can be found at `www.cs.york.ac.uk/~muehlber/blast/`. In our experience, Blast is rather difficult to apply by a practitioner during OS software development. This is because of (i) its limitations with respect to reasoning about pointers, (ii) several issues regarding usability, including bugs in `spec.opt`, and (iii) a lack of consistent documentation. Especially in the case of memory safety properties, massive changes to the source code were necessary which essentially requires one to know about a bug beforehand. However, it must be mentioned that Blast was not designed as a memory debugger. Indeed, Blast performed considerably better during our tests with locking properties; however, modifications on the source code were still necessary in most cases.

Blast performed nicely on the modified source code in our examples for locking properties. Even large portions of C code — up to 10k lines with several locks, status variables and a relatively complex program structure — were parsed and model checked within a few minutes on a modern PC. Hence, the techniques for abstraction and refinement as implemented in Blast are quite able to deal with most of the problems analysed in this paper. If its limitations are ironed out, Blast is likely to become a very usable and popular tool with OS software developers in the future.

Regarding future work we propose that our case study is repeated once the most problematic errors and restrictions in Blast are fixed. An analysis allowing one to draw *quantitative* conclusions concerning Blast's ability of finding certain programming problems could then give results that are more interesting to kernel developers. To this end, metrics for the evaluation of Blast are required, as is a more precise classification of the chosen examples.

Acknowledgements. We thank Radu Siminiceanu for his constructive comments and suggestions on a draft of this paper.

References

1. Ball, T. and Rajamani, S. K. Automatically validating temporal safety properties of interfaces. In *SPIN 2001*, vol. 2057 of *LNCS*, pp. 103–122.
2. Beyer, D., Chlipala, A. J., Henzinger, T. A., Jhala, R., and Majumdar, R. The BLAST query language for software verification. In *PEPM 2004*, pp. 201–202. ACM Press.
3. Beyer, D., Henzinger, T. A., Jhala, R., and Majumdar, R. Checking memory safety with BLAST. In *FASE 2005*, vol. 3442 of *LNCS*, pp. 2–18.
4. Breuer, P. T. and Pickin, S. Abstract interpretation meets model checking near the 10^6 LOC mark. In *AVIS 2006*. To appear in ENTCS.
5. Chaki, S., Clarke, E., Groce, A., Ouaknine, J., Strichman, O., and Yorav, K. Efficient verification of sequential and concurrent C programs. *FMSD*, 25(2–3):129–166, 2004.
6. Chou, A., Yang, J., Chelf, B., Hallem, S., and Engler, D. R. An empirical study of operating system errors. In *SOSP 2001*, pp. 73–88. ACM Press.
7. Clarke, E. M., Grumberg, O., and Peled, D. A. *Model checking*. MIT Press, 2000.
8. Corbet, J., Rubini, A., and Kroah-Hartmann, G. *Linux Device Drivers*. O'Reilly, 3rd edition, 2005.
9. Corbett et al, J. C. Bandera: Extracting finite-state models from Java source code. In *ICST 2000*, pp. 439–448. SQS Publishing.
10. Cousot, P. and Cousot, R. On abstraction in software verification. In *CAV 2002*, vol. 2404 of *LNCS*, pp. 37–56.
11. Engler, D. R., Chelf, B., Chou, A., and Hallem, S. Checking system rules using system-specific, programmer-written compiler extensions. In *OSDI 2000*. USENIX.
12. Engler, D. R. and Musuvathi, M. Static analysis versus software model checking for bug finding. In *VMCAI 2004*, vol. 2937 of *LNCS*, pp. 191–210.
13. Henzinger, T. A., Jhala, R., and Majumdar, R. Race checking by context inference. In *PLDI 2004*, pp. 1–13. ACM Press.
14. Henzinger, T. A., Jhala, R., Majumdar, R., and Sanvido, M. A. A. Extreme model cecking. In *Verification: Theory & practice*, vol. 2772 of *LNCS*, pp. 232–358, 2003.
15. Henzinger, T. A., Jhala, R., Majumdar, R., and Sutre, G. Lazy abstraction. In *POPL 2002*, pp. 58–70. ACM Press.
16. Henzinger et al, T. A. Temporal-safety proofs for systems code. In *CAV 2002*, vol. 2404 of *LNCS*, pp. 526–538.
17. Holzmann, G. J. *The SPIN model checker*. Addison-Wesley, 2003.
18. Jie, H. and Shivaji, S. Temporal safety verification of AVFS using BLAST. Project report, Univ. California at Santa Cruz, 2004.
19. Microsoft Corporation. Static driver verifier: Finding bugs in device drivers at compile-time. `www.microsoft.com/whdc/devtools/tools/SDV.mspx`.
20. Mong, W. S. Lazy abstraction on software model checking. Project report, Toronto Univ., Canada., 2004.
21. Necula, G. C., McPeaki, S., and Weimer, W. CCured: Type-safe retrofitting of legacy code. In *POPL 2002*, pp. 128–139. ACM Press.

A Examples Regarding Memory Safety

A.1 Checking Memory Safety: Example 1

Commit Overview

Commit Key `d2b0e84d195a341c1cc5b45ec2098ee23bc1fe9d`
Subject [PATCH] drivers/message/i2o/pci.c: fix a use-after-free
Description The Coverity checker spotted this obvious use-after-free
Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/message/i2o/pci.c
+++ b/drivers/message/i2o/pci.c
@@ -421,8 +421,8 @@ static int __devinit i2o_pci_probe(struc
          i2o_pci_free(c);
        free_controller:
-         i2o_iop_free(c);
          put_device(c->device.parent);
+         i2o_iop_free(c);
        disable:
          pci_disable_device(pdev);
```

Files

– `drivers/message/i2o/pci.c`

Comments

This is the running example used in Section 3, "Checking Memory Safety". The problem in this case is, that the pointer `c` is de-referenced in line 425 of the source file. However, the call of `i2o_iop_free(c)` in line 424 does nothing else than releasing `c`. De-referencing it afterwards results in undefined behaviour of the kernel. The bug has been fixed by simply swapping lines 424 and 425.

We have studied this example extensively using two different approaches. In the first place we used a temporal safety specification that ensures that the functions `i2o_iop_alloc()` (allocate memory for `c`), `put_device()` and `i2o_iop_free()` are called exactly in this sequence. Due to bugs in `spec.opt` this technique only worked for a manually simplified version of the source code under consideration.

In our second approach we modified the the source code of the driver in order to introduce the label `ERROR` by hand. Mainly, we added a status field to the struct `i2o_controller` and modified `i2o_iop_free()` not to release the pointer but to change this field. Further modifications to `put_device()` would then enable us to detect a wrong call order. However, BLAST failed in tracing content of our status field over the several function calls. In the paper we provide a rather simple example that shows BLAST's deficiencies in dealing with pointers.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=d2b0e84d195a341c1cc5b45ec2098ee23bc1fe9d`

A.2 Checking Memory Safety: Example 2

Commit Overview

Commit Key `2d6eac6c4fdaa69656d66c80754d267be233cc3f`
Subject [PATCH] drivers/infiniband/core/mad.c: fix a use-after-free
Description The Coverity checker spotted this obvious use-after-free caused by a wrong order of the cleanups.
Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/infiniband/core/mad.c
+++ b/drivers/infiniband/core/mad.c
@@ -356,9 +356,9 @@ error4:
          spin_unlock_irqrestore(&port_priv->reg_lock, flags);
          kfree(reg_req);
  error3:
-         kfree(mad_agent_priv);
- error2:
          ib_dereg_mr(mad_agent_priv->agent.mr);
+ error2:
+         kfree(mad_agent_priv);
  error1:
          return ret;
  }
```

Files

– `drivers/infiniband/core/mad.c`

Comments

This example is actually quite similar to Example 1. The bug results from a wrong oder of the labels used in the different error cases. If the execution of `ib_register_mad_agent()` ever jumps to either `error4` or `error3`, it will first release the pointer `mad_agent_priv` in line 359 but de-reference it again in line 361.
While experimenting with this error, we experienced the same problems as with Example 1.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=2d6eac6c4fdaa69656d66c80754d267be233cc3f`

A.3 Checking Memory Safety: Example 3

Commit Overview

Commit Key d0743a5b7b837334cb414b773529d51de3de0471
Subject [PATCH] drivers/scsi/dpt_i2o.c: fix a user-after-free
Description The Coverity checker spotted this obvious use-after-free
Requires Linux 2.6.14 kernel source as from git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git

---

```
--- a/drivers/scsi/dpt_i2o.c
+++ b/drivers/scsi/dpt_i2o.c
@@ -816,7 +816,7 @@ static int adpt_hba_reset(adpt_hba* pHba
 static void adpt_i2o_sys_shutdown(void)
 {
  adpt_hba *pHba, *pNext;
-struct adpt_i2o_post_wait_data *p1, *p2;
+ struct adpt_i2o_post_wait_data *p1, *old;

   printk(KERN_INFO"Shutting down Adaptec I2O controllers.\n");
   printk(KERN_INFO"   This could take a few minutes if there are ...
@@ -830,13 +830,14 @@ static void adpt_i2o_sys_shutdown(void)
  }

  /* Remove any timedout entries from the wait queue.  */
-p2 = NULL;
 // spin_lock_irqsave(&adpt_post_wait_lock, flags);
  /* Nothing should be outstanding at this point so just
   * free them
   */
-for(p1 = adpt_post_wait_queue; p1; p2 = p1, p1 = p2->next) {
-kfree(p1);
+ for(p1 = adpt_post_wait_queue; p1;) {
+ old = p1;
+ p1 = p1->next;
+ kfree(old);
  }
 // spin_unlock_irqrestore(&adpt_post_wait_lock, flags);
  adpt_post_wait_queue = NULL;
```

Files

– drivers/scsi/dpt‚Äòi2o.c

Comments

Despite being obvious, this programming error is quite difficult to verify. In some sense, the problem is similar to examples 1 and 2. In `adpt_i2o_sys_shutdown()` pointer is released using `kfree()` but de-referenced again afterwards. The bug resides in the `for`-loop in lines 838 to 840 of the source file. The loop is used to free a list of pointers to structs. Each of them contains a pointer `next` to the next element of the list. However, in the first cycle the loop stores the first element of this list in pointer `p1`, checks whether it does not equal `NULL`, makes `p1` point to the next element and frees `p1`. In the next cycle it checks whether `p1` still does not equal `NULL` and sets it to `p1->next`. This loop behaviour actually contains two errors: Firstly, the second element of the list is freed without checking whether it might equal `NULL`. Secondly, the already freed second element is de-referenced again in the second loop cycle.
Verifying a linked list package has already been attempted by Mong as referenced in our paper. We basically experienced similar shortcomings as he did. The major issue is, that BLAST is not able to keep track of a set of pointers. Probably some sort of a heap model would be required in order to make BLAST able to detect problems like this.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=d0743a5b7b837334cb414b773529d51de3de0471`

A.4 Checking Memory Safety: Example 4

Commit Overview

Commit Key	`6968ecfca8822055cfe121214c0786e4eecc038e`
Subject	[PATCH] apci: fix NULL deref in video/lcd/brightness
Description	Fix Null pointer deref in video/lcd/brightness `http://bugzilla.kernel.org/show_bug.cgi?id=5571`
Requires	Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/acpi/video.c
+++ b/drivers/acpi/video.c
@@ -813,7 +813,7 @@ acpi_video_device_write_brightness(struc

  ACPI_FUNCTION_TRACE("acpi_video_device_write_brightness");

-if (!dev || count + 1 > sizeof str)
+ if (!dev || !dev->brightness || count + 1 > sizeof str)
  return_VALUE(-EINVAL);

  if (copy_from_user(str, buffer, count))
```

Files

- `drivers/acpi/video.c`

Comments

This is a classical example for a function that does not properly check whether its parameters are valid. While line 816 of `acpi_video_device_write_brightness()` contains a test ensuring that `file->private_data->private` does not equal `NULL`, there is no such test for the component `file->private_data->private->brightness`, which is de-referenced in line 829 of the listing.
Finding this problem using BLAST is rather difficult since BLAST does not provide a way to specify that "whenever a pointer is de-referenced, it must not equal NULL". In "Checking Memory Safety with Blast" by Beyer et al. the problem is addressed by automatically inserting runtime tests into the source code under consideration and then using BLAST to check whether the newly introduced code is reachable. However, our case is more difficult since `acpi_video_device_write_brightness()` is not called directly but via a function pointer assigned in line 939 of the source file.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=6968ecfca8822055cfe121214c0786e4eecc038e`

A.5 Checking Memory Safety: Example 5

Commit Overview

Commit Key `abd559b1052e28d8b9c28aabde241f18fa89090b`

Subject [PATCH] sbp2: fix deadlocks and delays on device removal/rmmod

Description Fixes for deadlocks of the ieee1394 and scsi subsystems and long delays in futile error recovery attempts when SBP-2 devices are removed or drivers are unloaded.

- Complete commands quickly with DID_NO_CONNECT if the 1394 node is gone or if the 1394 low-level driver was unloaded.
- Skip unnecessary work in the eh_abort_handler and eh_device_reset_handler if the node or 1394 low-level driver is gone.
- Let scsi's high-level shut down gracefully when sbp2 is being unloaded or detached from the 1394 unit. A call to scsi_remove_device is added for this purpose, which requires us to store a scsi_device pointer.
- scsi_device pointer is obtained from slave_alloc hook and cleared by slave_destroy. This avoids usage of the pointer after the scsi device was deleted e.g. by the user via scsi_mod's sysfs interface.

Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/ieee1394/sbp2.c
+++ b/drivers/ieee1394/sbp2.c
@@ -596,6 +596,14 @@ static void sbp2util_mark_command_comple
  spin_unlock_irqrestore(&scsi_id->sbp2_command_orb_lock, flags);
 }

+/*
+ * Is scsi_id valid? Is the 1394 node still present?
+ */
+static inline int sbp2util_node_is_available(struct scsi_id_instance_...
+{
+ return scsi_id && scsi_id->ne && !scsi_id->ne->in_limbo;
+}

[This diff would be several pages long and has been shortened.]
```

Files

– `drivers/ieee1394/sbp2.c`

Comments

Regarding memory safety, the interesting error in this example is related to the use of the `scsi_device` pointer, which might be used after removing the related device and freeing the pointer. This can happen due to concurrent interaction of a secondary kernel thread. Since it does not seem to be possible to force BLAST to consider this behaviour, it was not possible to detect the error.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=abd559b1052e28d8b9c28aabde241f18fa89090b`

A.6 Checking Memory Safety: Example 6

Commit Overview

Commit Key `3fd1bb9baa394856b112e5edbfd3893d92dd1149`
Subject [PATCH] hwmon: Off-by-one error in fscpos driver
Description Coverity uncovered an off-by-one error in the fscpos driver, in function set_temp_reset(). Writing to the temp3_reset sysfs file will lead to an array overrun, in turn causing an I2C write to a random register of the FSC Poseidon chip. Additionally, writing to temp1_reset and temp2_reset will not work as expected. The fix is straightforward.
Requires Linux 2.6.13 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.13.y.git`

```
--- a/drivers/hwmon/fscpos.c
+++ b/drivers/hwmon/fscpos.c
@@ -167,7 +167,7 @@ static ssize_t set_temp_reset(struct i2c
  "experience to the module author.\n");

  /* Supported value: 2 (clears the status) */
-fscpos_write_value(client, FSCPOS_REG_TEMP_STATE[nr], 2);
+ fscpos_write_value(client, FSCPOS_REG_TEMP_STATE[nr - 1], 2);
  return count;
 }
```

Files

– `drivers/hwmon/fscpos.c`

Comments

Despite the simple patch, this bug is not easily understood due to the structure of the source file. The function `set_temp_reset()` operates on the array `FSCPOS_REG_TEMP_STATE`, containing three values. Therefore, calls of `set_temp_reset()` must have the parameter `nr` be in the range of 0 to 2. Unfortunately, there are no such calls visible since they are generated during macro expansion by the preprocessor. To fully understand the bug you may want to look at preprocessed code.
This bug can be easily found using BLAST by introducing an an additional check for the value of the `nr` argument passed to `set_temp_reset()` in the preprocessed source file. Fully automatic discovery of the bug seems to be impossible.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.13.y.git;a=commit;h=3fd1bb9baa394856b112e5edbfd3893d92dd1149`

A.7 Checking Memory Safety: Example 7

Commit Overview

Commit Key 703b69791369263e1d15f88f3e6aed02c1514fc2
Subject [PATCH] Fix another crash in ip_nat_pptp (CVE-2006-0037)
Description The PPTP NAT helper calculates the offset at which the packet needs to be mangled as difference between two pointers to the header. With non-linear skbs however the pointers may point to two seperate buffers on the stack and the calculation results in a wrong offset beeing used.
Requires Linux 2.6.14 kernel source as from git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git

---

```
--- a/net/ipv4/netfilter/ip_nat_helper_pptp.c
+++ b/net/ipv4/netfilter/ip_nat_helper_pptp.c
@@ -148,14 +148,14 @@ pptp_outbound_pkt(struct sk_buff **pskb,
 {
  struct ip_ct_pptp_master *ct_pptp_info = &ct->help.ct_pptp_info;
  struct ip_nat_pptp *nat_pptp_info = &ct->nat.help.nat_pptp_info;
-
-u_int16_t msg, *cid = NULL, new_callid;
+ u_int16_t msg, new_callid;
+ unsigned int cid_off;

 [This diff would be several pages long and has been shortened.]
```

Files

– net/ipv4/netfilter/ip_nat_helper_pptp.c

Comments

This bug is nicely explained above. It results from a bogous assumption made before computing a pointer. Pointer arithmetic is rather difficult to analyse and to verify in general. One way to deal with problems like this would be to do the verification in respect of a memory model. However, this bug is out of scope for BLAST.

Source: http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=703b69791369263e1d15f88f3e6aed02c1514fc2

A.8 Checking Memory Safety: Example 8

Commit Overview

Commit Key 67a69cdd748de32d9991056c207f7ab3798230a5
Subject [PATCH] PCI: fix hotplug double free
Description With the brackets missed out func could be freed twice. Found by Coverity tool
Requires Linux 2.6.1 kernel source as from git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.11.y.git

---

```
--- a/drivers/pci/hotplug/pciehp_ctrl.c
+++ b/drivers/pci/hotplug/pciehp_ctrl.c
@@ -1354,10 +1354,11 @@ static u32 remove_board(struct pci_func
  dbg("PCI Bridge Hot-Remove s:b:d:f(%02x:%02x:%02x:%02x)\n",
  ctrl->seg, func->bus, func->device, func->function);
  bridge_slot_remove(func);
-} else
+ } else {
  dbg("PCI Function Hot-Remove s:b:d:f(%02x:%02x:%02x:%02x)\n",
  ctrl->seg, func->bus, func->device, func->function);
  slot_remove(func);
+ }

  func = pciehp_slot_find(ctrl->slot_bus, device, 0);
  }
```

Files

– drivers/pci/hotplug/pciehp_ctrl.c

Comments

The error in this example resides in lines 1357 to 1360 of the source file. The two functions bridge_slot_remove() and slot_remove() basically do the same thing – calling kfree() on the parameter. Hence, without the brackets around lines 1358 to 1360 the pointer func could be freed twice. Despite this, slot_remove() performs several checks on func whereas the already freed pointer gets de-referenced.
Using BLAST we were able to find this bug on a manually simplified version of the source code.

Source: http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.11.y.git;a=commit;h=67a69cdd748de32d9991056c207f7ab3798230a5

B Examples Regarding Locking Properties

B.1 Checking Locking Properties: Example 1

Commit Overview

Commit Key `d7283d61302798c0c57118e53d7732bec94f8d42`
Subject [PATCH] libata: locking rewrite (== fix)
Description [libata] locking rewrite (== fix)
A lot of power packed into a little patch.
This change eliminates the sharing between our controller-wide spinlock and the SCSI core's Scsi_Host lock. As the locking in libata was already highly compartmentalized, always referencing our own lock, and never scsi_host::host_lock.
As a side effect, this change eliminates a deadlock from calling scsi_finish_command() while inside our spinlock.
Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/scsi/libata-core.c
+++ b/drivers/scsi/libata-core.c
@@ -3916,8 +3916,6 @@ static void ata_host_init(struct ata_por
  host->unique_id = ata_unique_id++;
  host->max_cmd_len = 12;

-scsi_assign_lock(host, &host_set->lock);
-
  ap->flags = ATA_FLAG_PORT_DISABLED;
  ap->id = host->unique_id;
  ap->host = host;
--- a/drivers/scsi/libata-scsi.c
+++ b/drivers/scsi/libata-scsi.c
@@ -39,6 +39,7 @@
 #include <scsi/scsi.h>
 #include "scsi.h"
 #include <scsi/scsi_host.h>
+#include <scsi/scsi_device.h>
 #include <linux/libata.h>
 #include <asm/uaccess.h>

@@ -1565,8 +1566,12 @@ int ata_scsi_queuecmd(struct scsi_cmnd *
  struct ata_port *ap;
  struct ata_device *dev;
  struct scsi_device *scsidev = cmd->device;
+ struct Scsi_Host *shost = scsidev->host;

-ap = (struct ata_port *) &scsidev->host->hostdata[0];
```

```
+ ap = (struct ata_port *) &shost->hostdata[0];
+
+ spin_unlock(shost->host_lock);
+ spin_lock(&ap->host_set->lock);

  ata_scsi_dump_cdb(ap, cmd);

@@ -1589,6 +1594,8 @@ int ata_scsi_queuecmd(struct scsi_cmnd *
  ata_scsi_translate(ap, dev, cmd, done, atapi_xlat);

 out_unlock:
+ spin_unlock(&ap->host_set->lock);
+ spin_lock(shost->host_lock);
  return 0;
 }
```

Files

- `drivers/scsi/scsi.c`
- `drivers/scsi/scsi_lib.c`

Comments

This is the running example for Section 4, "Checking Locking Properties". The example code contains a deadlock caused by a API rule violation: The functions `spin_lock()` and `spin_unlock()` must be called alternating on a specific lock. However, the problem is not obvious. Firstly, one has to know that whenever the program execution enters the function `ata_scsi_queuecmd()`, a lock on `shost->host_lock` is held. In an unlikely case the `if`-statement in line 1574 of `libata-scsi.c` will evaluate to true and line 1576:`done(cmd);` will be executed. This `done()` is a function pointer pointing to `scsi_finish_command()` (`drivers/scsi/scsi.c`), which will call 876:`scsi_device_unbusy()`. In line 447 of `drivers/scsi/scsi_lib.c` we see that this function will again lock `shost->host_lock`. Deadlock.

In order to find this bug using BLAST, simplifications needed to be applied to `libata-scsi.c`. We simplified several data structures, removed the use of function pointers and put all functions required into one C source file. The verification was then done using the temporal safety specification as given in the paper.

However, it was much more difficult to prove the patched version of the code to be free of this deadlock. Since it involves pointers to two different spinlocks; and pointers are not reasonably tracked by BLAST, we had to rewrite parts of the spinlock-API and used two global integer variables to represent the state of each lock.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=d7283d61302798c0c57118e53d7732bec94f8d42`

B.2 Checking Locking Properties: Example 2

Commit Overview

Commit Key `fe2e17a405a58ec8a7138fee4ebe101858b636e0`
Subject [PATCH] dpt_i2o fix for deadlock condition
Description Miquel van Smoorenburg <`miquels@cistron.nl`> forwarded me this fix to resolve a deadlock condition that occurs due to the API change in 2.6.13+ kernels dropping the host locking when entering the error handling. They all end up calling adpt_i2o_post_wait(), which if you call it unlocked, might return with host_lock locked anyway and that causes a deadlock.
Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
-- a/drivers/scsi/dpt_i2o.c
+++ b/drivers/scsi/dpt_i2o.c
@@ -660,7 +660,12 @@ static int adpt_abort(struct scsi_cmnd *
  msg[2] = 0;
  msg[3]= 0;
  msg[4] = (u32)cmd;
-if( (rcode = adpt_i2o_post_wait(pHba, msg, sizeof(msg), FOREVER))...
+ if (pHba->host)
+ spin_lock_irq(pHba->host->host_lock);
+ rcode = adpt_i2o_post_wait(pHba, msg, sizeof(msg), FOREVER);
+ if (pHba->host)
+ spin_unlock_irq(pHba->host->host_lock);
+ if (rcode != 0) {
  if(rcode == -EOPNOTSUPP ){

[This diff would be several pages long and has been shortened.]
```

Files

– `drivers/scsi/dpt_i2o.c`

Comments

The problem in this case study is well explained above and the fixes are straightforward. We only want to add, that the trouble-causing call of `spin_lock_irq()` on the host lock can be found in line 1172 of the source file provided.
Tracing this error shaped up as being straightforward, too. However, the use of pointers and complex data structures (i.e. `pHba->host->host_lock`) forced us to use simple integer variables instead of the `spinlock_t` pointers.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=fe2e17a405a58ec8a7138fee4ebe101858b636e0`

B.3 Checking Locking Properties: Example 3

Commit Overview

Commit Key `910573c7c4aced8fd5f45c334cc67862e3424d92`

Subject [PATCH] ieee1394/sbp2: fixes for hot-unplug and module unloading

Description Fixes for reference counting problems, deadlocks, and delays when SBP-2 devices are unplugged or unbound from sbp2, or when unloading of sbp2/ ohci1394/ pcilynx is attempted.

Most often reported symptoms were hotplugs remaining undetected once a FireWire disk was unplugged since the knodemgrd kernel thread went to uninterruptible sleep, and "modprobe -r sbp2" being unable to complete because still being in use.

Patch is equivalent to commit abd559b1052e28d8b9c28aabde241f18fa89090b in 2.6.14-rc3 plus a fix which is necessary together with 2.6.13's scsi core API (linux1394.org commit r1308 by Ben Collins).

Requires Linux 2.6.13 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.13.y.git`

```
--- a/drivers/ieee1394/sbp2.c
+++ b/drivers/ieee1394/sbp2.c
@@ -596,6 +596,11 @@ static void sbp2util_mark_command_comple
  spin_unlock_irqrestore(&scsi_id->sbp2_command_orb_lock, flags);
 }

+static inline int sbp2util_node_is_available(struct scsi_id_instance_...
+{
+ return scsi_id && scsi_id->ne && !scsi_id->ne->in_limbo;
+}
+

 /**************************************
@@ -631,11 +636,23 @@ static int sbp2_remove(struct device *de
 {
  struct unit_directory *ud;
  struct scsi_id_instance_data *scsi_id;
+ struct scsi_device *sdev;

  SBP2_DEBUG("sbp2_remove");

[This diff would be several pages long and has been shortened.]
```

Files

– `drivers/ieee1394/sbp2.c`

Comments

See Example 4.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.13.y.git;a=commit;h=910573c7c4aced8fd5f45c334cc67862e3424d92`

B.4 Checking Locking Properties: Example 4

Commit Overview

Commit Key `abd559b1052e28d8b9c28aabde241f18fa89090b`

Subject [PATCH] sbp2: fix deadlocks and delays on device removal/rmmod

Description Fixes for deadlocks of the ieee1394 and scsi subsystems and long delays in futile error recovery attempts when SBP-2 devices are removed or drivers are unloaded.

- Complete commands quickly with DID_NO_CONNECT if the 1394 node is gone or if the 1394 low-level driver was unloaded.
- Skip unnecessary work in the eh_abort_handler and eh_device_reset_handler if the node or 1394 low-level driver is gone.
- Let scsi's high-level shut down gracefully when sbp2 is being unloaded or detached from the 1394 unit. A call to scsi_remove_device is added for this purpose, which requires us to store a scsi_device pointer.
- scsi_device pointer is obtained from slave_alloc hook and cleared by slave_destroy. This avoids usage of the pointer after the scsi device was deleted e.g. by the user via scsi_mod's sysfs interface.

Requires Linux 2.6.14 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.14.y.git`

```
--- a/drivers/ieee1394/sbp2.c
+++ b/drivers/ieee1394/sbp2.c
@@ -596,6 +596,14 @@ static void sbp2util_mark_command_comple
  spin_unlock_irqrestore(&scsi_id->sbp2_command_orb_lock, flags);
 }

+/*
+ * Is scsi_id valid? Is the 1394 node still present?
+ */
+static inline int sbp2util_node_is_available(struct scsi_id_instance_...
+{
+ return scsi_id && scsi_id->ne && !scsi_id->ne->in_limbo;
```

```
+}
+

 /*****************************************
@@ -631,11 +639,23 @@ static int sbp2_remove(struct device *de
 {
  struct unit_directory *ud;
  struct scsi_id_instance_data *scsi_id;
+ struct scsi_device *sdev;

[This diff would be several pages long and has been shortened.]
```

Files

– `drivers/ieee1394/sbp2.c`

Comments

This is an interesting case in which no detailed explanation of the deadlock that should be found in this source file is given. Despite trying several different attempts to model check simplifications of this source code, we did not find any locking related problems, either in the version of the source file used in Linux 2.6.13 nor in 2.6.14.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.14.y.git;a=commit;h=abd559b1052e28d8b9c28aabde241f18fa89090b`

B.5 Checking Locking Properties: Example 5

Commit Overview

Commit Key `3515d0161d55d2fa1a340932625f94240a68c262`
Subject [PATCH] late spinlock initialization in ieee1394/ohci
Description spinlock used in irq handler should be initialized before registering irq, even if we know that our device has interrupts disabled; handler is registered shared and taking spinlock is done unconditionally. As it is, we can and do get oopsen on boot for some configuration, depending on irq routing - I've got a reproducer.
Requires Linux 2.6.13 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.13.y.git`

```
--- a/drivers/ieee1394/ohci1394.c
+++ b/drivers/ieee1394/ohci1394.c
@@ -478,7 +478,6 @@ static void ohci_initialize(struct ti_oh
  int num_ports, i;
```

```
   spin_lock_init(&ohci->phy_reg_lock);
 -spin_lock_init(&ohci->event_lock);

   /* Put some defaults to these undefined bus options */
   buf = reg_read(ohci, OHCI1394_BusOptions);
 @@ -3402,7 +3401,14 @@ static int __devinit ohci1394_pci_probe(
   /* We hopefully don't have to pre-allocate IT DMA like we did
    * for IR DMA above. Allocate it on-demand and mark inactive. */
   ohci->it_legacy_context.ohci = NULL;
 + spin_lock_init(&ohci->event_lock);

 + /*
 +  * interrupts are disabled, all right, but... due to SA_SHIRQ we
 +  * might get called anyway.  We'll see no event, of course, but
 +  * we need to get to that "no event", so enough should be initialized
 +  * by that point.
 +  */
   if (request_irq(dev->irq, ohci_irq_handler, SA_SHIRQ,
    OHCI1394_DRIVER_NAME, ohci))
   FAIL(-ENOMEM, "Failed to allocate shared interrupt %d", dev->irq);
```

Files

– `drivers/ieee1394/ohci1394.c`

Comments

In this case, a spinlock might be used uninitialised, which represents a violation of the usage rules of the spinlock API. However, this violation can only occur in an environment where several kernel threads are running concurrently and may interfere the initialisation of a device. Modelling this behaviour turned out to be impossible.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.13.y.git;a=commit;h=3515d0161d55d2fa1a340932625f94240a68c262`

B.6 Checking Locking Properties: Example 6

Commit Overview

Commit Key `f7cfcc72b365dc62cd01e1920f3f0b4e053f7735`
Subject [PATCH] Fix deadlock in br_stp_disable_bridge
Description Looks like somebody forgot to use the _bh spin_lock variant. We ran into a deadlock where br->hello_timer expired while br_stp_disable_br() walked br->port_list.
Requires Linux 2.6.15 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.15.y.git`

```
--- a/net/bridge/br_stp_if.c
+++ b/net/bridge/br_stp_if.c
@@ -67,7 +67,7 @@ void br_stp_disable_bridge(struct net_br
 {
  struct net_bridge_port *p;

-spin_lock(&br->lock);
+ spin_lock_bh(&br->lock);
  list_for_each_entry(p, &br->port_list, list) {
  if (p->state != BR_STATE_DISABLED)
  br_stp_disable_port(p);
@@ -76,7 +76,7 @@ void br_stp_disable_bridge(struct net_br

  br->topology_change = 0;
  br->topology_change_detected = 0;
-spin_unlock(&br->lock);
+ spin_unlock_bh(&br->lock);

  del_timer_sync(&br->hello_timer);
  del_timer_sync(&br->topology_change_timer);
```

Files

– `net/bridge/br_stp_if.c`

Comments

This API usage rule violation is rather obvious and the error was easily detected using BLAST.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.15.y.git;a=commit;h=f7cfcc72b365dc62cd01e1920f3f0b4e053f7735`

B.7 Checking Locking Properties: Example 7

Commit Overview

Commit Key `8fef8ea2a1f28a7611ad0b8ff7b48ceb38db9535`

Subject [PATCH] fix deadlock in ext2

Description Fix a deadlock possible in the ext2 file system implementation. This deadlock occurs when a file is removed from an ext2 file system which was mounted with the "sync" mount option.

The problem is that ext2_xattr_delete_inode() was invoking the routine, sync_dirty_buffer(), using a buffer head which was previously locked via lock_buffer(). The first thing that sync_dirty_buffer() does is to lock the buffer head that it was passed. It does this via lock_buffer(). Oops.

The solution is to unlock the buffer head in ext2_xattr_delete_inode() before invoking sync_dirty_buffer(). This makes the code in ext2_xattr_delete_inode() obey the same locking rules as all other callers of sync_dirty_buffer() in the ext2 file system implementation.

Requires Linux 2.6.15 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.15.y.git`

```
--- a/fs/ext2/xattr.c
+++ b/fs/ext2/xattr.c
@@ -796,18 +796,20 @@ ext2_xattr_delete_inode(struct inode *in
  ext2_free_blocks(inode, EXT2_I(inode)->i_file_acl, 1);
  get_bh(bh);
  bforget(bh);
+ unlock_buffer(bh);
  } else {
  HDR(bh)->h_refcount = cpu_to_le32(
  le32_to_cpu(HDR(bh)->h_refcount) - 1);
  if (ce)
  mb_cache_entry_release(ce);
+ ea_bdebug(bh, "refcount now=%d",
+ le32_to_cpu(HDR(bh)->h_refcount));
+ unlock_buffer(bh);
  mark_buffer_dirty(bh);
  if (IS_SYNC(inode))
  sync_dirty_buffer(bh);
  DQUOT_FREE_BLOCK(inode, 1);
  }
-ea_bdebug(bh, "refcount now=%d", le32_to_cpu(HDR(bh)->h_refcount) - 1);
-unlock_buffer(bh);
  EXT2_I(inode)->i_file_acl = 0;
```

Files

– `fs/ext2/xattr.c`

Comments

This locking problem involves a locking API very similar to spinlocks. As the error is well explained by the developer who submitted the patch, it was detected with BLAST after a few simplifications to the source code.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.15.y.git;a=commit;h=8fef8ea2a1f28a7611ad0b8ff7b48ceb38db9535`

B.8 Checking Locking Properties: Example 8

Commit Overview

Commit Key `fa0726854c4f03f9d2d1e00bb3b67a49ce490c32`

Subject [PATCH] ext3: fix race between ext3 make block reservation and reservation window ...

Description This patch fixed a race between ext3_discard_reservation() and ext3_try_to_allocate_with_rsv().
There is a window where ext3_discard_reservation will remove an already unlinked reservation window node from the filesystem reservation tree: It thinks the reservation is still linked in the filesystem reservation tree, but it is actually temperately removed from the tree by allocate_new_reservation() when it failed to make a new reservation from the current group and try to make a new reservation from next block group.
Here is how it could happen:

```
CPU 1
try to allocate a block in group1 with given
 reservation window my_rsv
ext3_try_to_allocate_with_rsv(group
----copy reservation window my_rsv into local
    rsv_copy
ext3_try_to_allocate(...rsv_copy)
----no free block in existing reservation window,
----need a new reservation window
spin_lock(&rsv_lock);

CPU 2
ext3_discard_reservation
if (!rsv_is_empty()
----this is true
```

```
spin_lock(&rsv_lock)
----waiting for thread 1

CPU 1:
allocate_new_reservation
failed to reserve blocks in this group
remove the window from the tree
rsv_window_remove(my_rsv)
----window node is unlinked from the tree here
return -1
spin_unlock(&rsv_lock)
ext3_try_to_allocate_with_rsv() failed in this
 group
group++

CPU 2
spin_lock(&rsv_lock) succeed
rsv_remove_window ()
---------------break, trying to remove a unlinked
 node from the tree
....

CPU 1:
ext3_try_to_allocate_with_rsv(group, my_rsv)
rsv_is_empty is true, need a new reservation window
spin_lock(&rsv_lock);
---------------spinning forever
```

We need to re-check whether the reservation window is still linked to the tree after grab the rsv_lock spin lock in ext3_discard_reservation, to prevent panic in rsv_remove_window->rb_erase.

Requires Linux 2.6.11 kernel source as from `git://git.kernel.org/pub/scm/linux/kernel/git/gregkh/linux-2.6.11.y.git`

```
--- a/fs/ext3/balloc.c
+++ b/fs/ext3/balloc.c
@@ -268,7 +268,8 @@ void ext3_discard_reservation(struct ino

  if (!rsv_is_empty(&rsv->rsv_window)) {
  spin_lock(rsv_lock);
-rsv_window_remove(inode->i_sb, rsv);
+ if (!rsv_is_empty(&rsv->rsv_window))
+ rsv_window_remove(inode->i_sb, rsv);
  spin_unlock(rsv_lock);
  }
 }
```

Files

– `fs/ext3/balloc.c`

Comments

Although this problem is extremely well explained, it was not possible to detect it using BLAST. It requires two kernel threads to operate concurrently on the driver involved, which cannot be modelled in BLAST.

Source: `http://www.kernel.org/git/?p=linux/kernel/git/stable/linux-2.6.11.y.git;a=commit;h=fa0726854c4f03f9d2d1e00bb3b67a49ce490c32`