

We politely ask for 2 grace periods for this project and apologize for any inconvenience caused by our delay.

CS510: Project2

Chufeng Gao gao513@purdue.edu

Meng-Chieh Lin lin1055@purdue.edu

Part 1 (b). Many bugs generated by our detection tool turn out to be false positives. The first reason is that the default algorithm does not expand scopes and check the partners in them during the detecting process, which results in many misclassified to appm ear wseeithout some of their "partners". But in fact, they are paired.

The second reason is that some of the reported "bugs" are intentionally designed by developers for some purposes, meaning those "bugs" should be treated as normal practices.

The pairs we want to discuss are (*apr_hash_make*, *apr_hash_set*) and (*apr_array_make*, *apr_array_push*). For the first pair, one of corresponding bug reports is below

```
bug: apr_hash_make in process_command_config, pair: (apr_hash_make, apr_hash_set), support: 8,
```

which indicates that *apr_hash_make* is unpaired in function *process_command_config*. However, when expanding the functions called in *process_command_config*, we find that function *apr_hash_set* is called in function *ap_add_module_commands* which is in the function *rebuild_conf_hash* called in *process_command_config*. In other words, *apr_hash_make* is paired with *apr_hash_set* if the default algorithm expands functions into second level. For the second pair, the one of corresponding bug reports is below

```
bug: apr_array_push in ap_directory_walk, pair: (apr_array_make, apr_array_push), support: 84, confide
```

which implies that function *apr_array_push* is unpaired in function *ap_directory_walk*. However, if we expand *ap_directory_walk*, it is not difficult to find that function *prep_walk_cache*, which is called in *ap_directory_walk*, calls function *apr_array_make* in its body. Therefore, we consider this bug report as a false positive with the same reason.

Part 1 (c). Our algorithms applies the concept, Inter-Procedural Analysis, to expand scopes when detection bugs with a given level parameter. To avoid misunderstanding, we call the functions, including scopes, in the example listed in page 2 of project2 functions here.

There are two process when we can perform function inline(expansion of scopes). One is the process when we are counting supports and calculating confidences. The other is the process when we are detecting bugs using the pairing assumption inferred from the counted supports and confidences. If we perform function inline during the counting process, the result could be unpredictable since the expansion process both increases method

support and pair support. Therefore, we choose only to expand scopes during the detection process.

We develop two algorithms and compare them with default one (we use the algorithm give in part a as a baseline). After calculating all supports and confidences and generating a map which contains pairs that should appear, our first algorithm detects bugs by expanding any functions that involve other functions while reserving those expanded functions. Different from our first algorithm, the second one expands those functions without keeping them while detecting bugs.

Take `scopeX{A(); B(); C(); D();}` as an example, if we expand it in a non-reservation fashion, it will look as `A(); B(); C(); D();`. If we expand it in a reservation fashion, it will look as `{scopeX(); A(); B(); C(); D();}`

Below are the experiments we examine¹ on three algorithms: likely-invariants (*LI*), expansion-involved likely-invariants with scope (*EILIS*), and expansion-involved likely-invariants without scope (*EILI*).

The following figures always show the result of the three algorithms from left to right in the order of *EILI*(Remove scope after expansion), *LI*(Baseline, no expansion performed), and *EILIS*(Reserve scope after expansion). The level in following figures are just for *EILI* and *EILIS*.

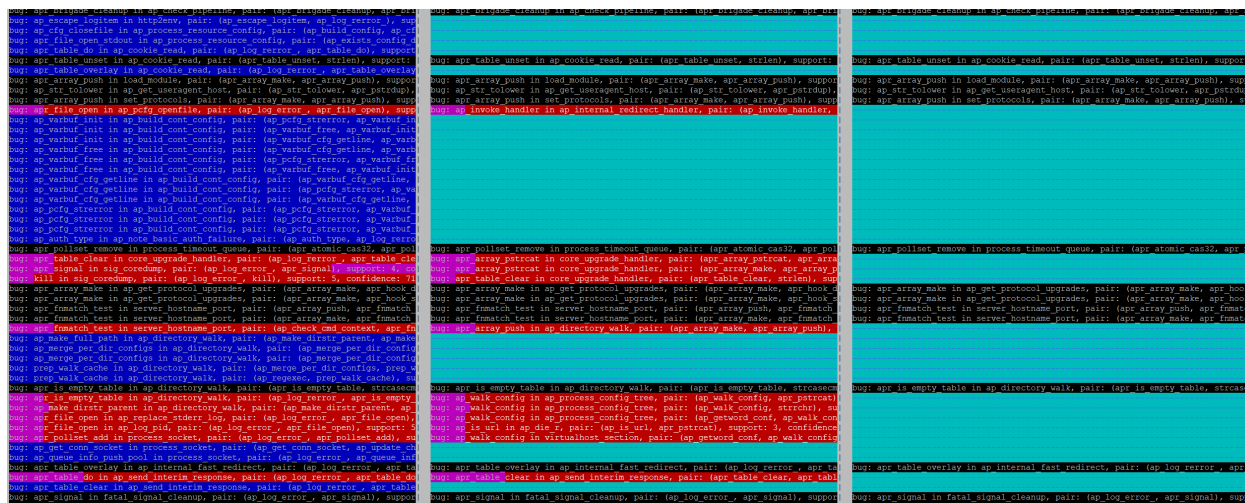


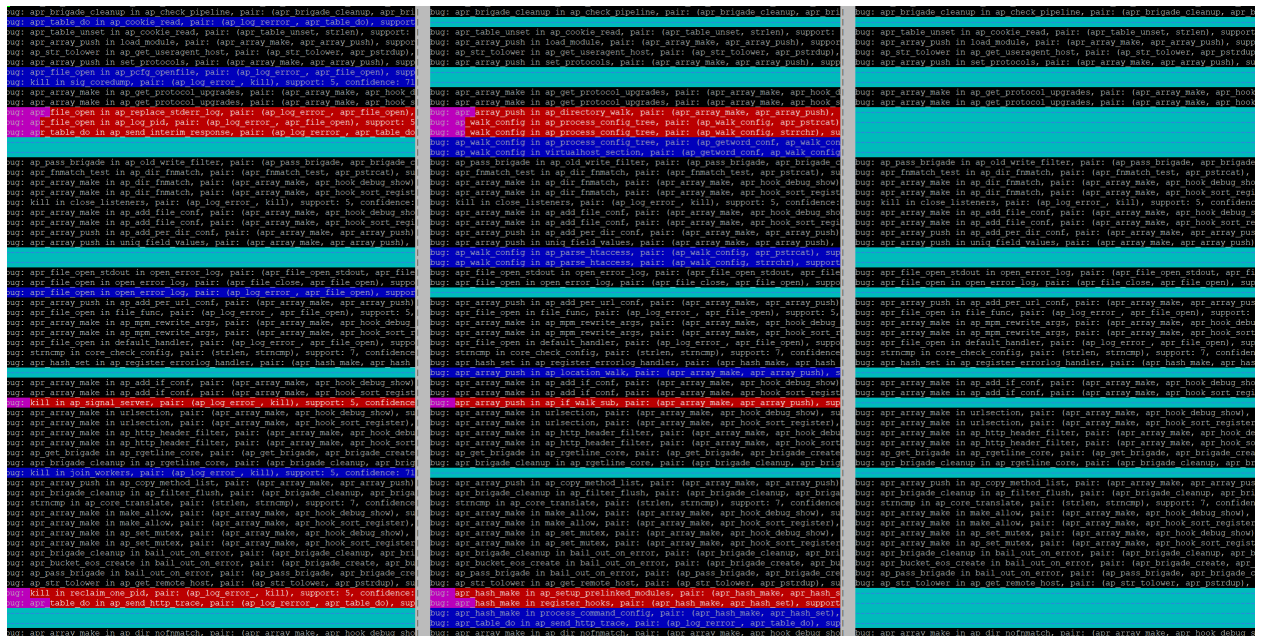
Figure 1: Support = 3, Confidence = 65, Level = 2

In Figure 1, we can see that *EILI*(Expansion with scopes no reserved) performs even worse than *LI*. Because the expanded functions do not remain after expansion, pairs that contain those expanded scopes are incorrectly consider as a bug. For *EILIS*(Expansion with scopes reserved), it behaves better than other two algorithms. The reason is that it digs deep into nested functions to find any possible contained pairs, which leads to better results in the end. Obviously, the assumption can be proved from the number of detected

¹To run our algorithm, just add an argument, level, after argument T_CONFIDENCE

[illegible]

In Figure 2, the bugs detected by each algorithm are 436, 253, and 180 respectively. As the level of expansion increases, the number of false positives decreases.



3

In Figure 3, the bugs detected by each algorithm are 117, 118, and 101 respectively.

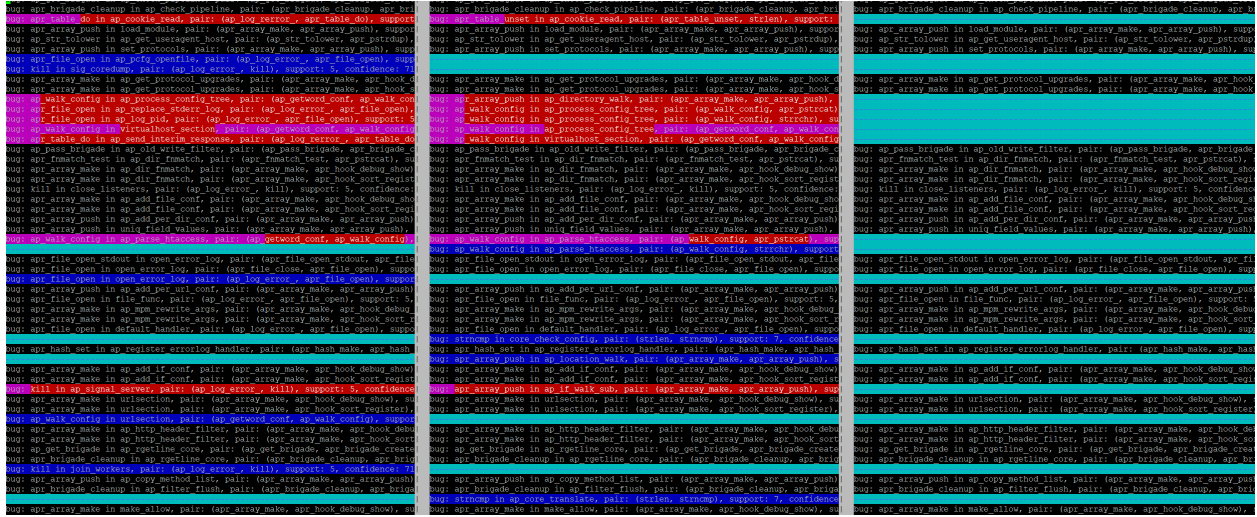


Figure 4: Support = 5, Confidence = 70, Level = 5

In Figure 4, the bugs detected by each algorithm are 119, 118, and 95 respectively. In conclusion, our first algorithm, *EILIS*, performs the best among the three in eliminating false positives. By the way, we also try the algorithm of expanding functions while counting supports and confidences, but the result varies in a huge range. Therefore, we no more provide any experiment results here.

Part 1 (d). Since there is little ground truth data of project 2, we choose not to refer to any published paper for other potential solutions. Instead, we derive our own way to reduce the false positives. To further enhance the algorithm introduced in part1(c), we create an algorithm, *EILISP*. Based on *EILIS*, we analyze the “partner list” we acquired for each function after counting the supports and comparing supports and confidences to the corresponding thresholds. We use the “partner list” to help determine whether the function call is a bug by calculating the ratio of missing partners.

Take the following case as an example.

function	partners	void K(){ A(); B(); C(); D(); }
A	B,C,D,E	

After building the partner list for function A, we calculate the ratio mentioned above after expanding functions. The ratio in this case is 1/4 since only E does not appear in node K. If the achieved ratio is lower than the unpaired ratio threshold passed as an argument into the program, we assume the A appearing without B is not a bug.

The intuition here is that when A appears without B, the cause could be that the function B is missing here instead of function A is redundant. If most partners of A appear

in this function node, we can assume there should be a function call of A here. Thus the reason of A unpaired with B here is that the develop forgets to call B in the function node.

The results of algorithms showed below are always in the order as: *LI*(Baseline), *EILIS*(Expansion with scope reserved), and *EILISP*(Expansion with scope reserved and enhanced with unpaired partner ratio). The ratio argument is only applied on *EILISP*.

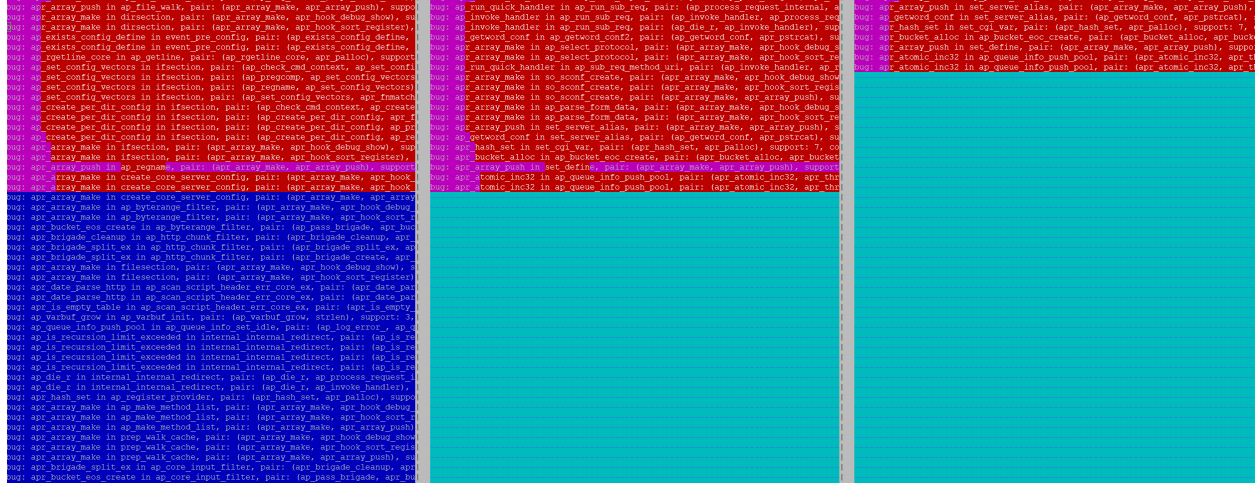


Figure 5: Support = 3, Confidence = 65, Level = 2, ratio = 40

In Figure 5, the bugs detected by each algorithm are 253, 200, and 188 respectively.

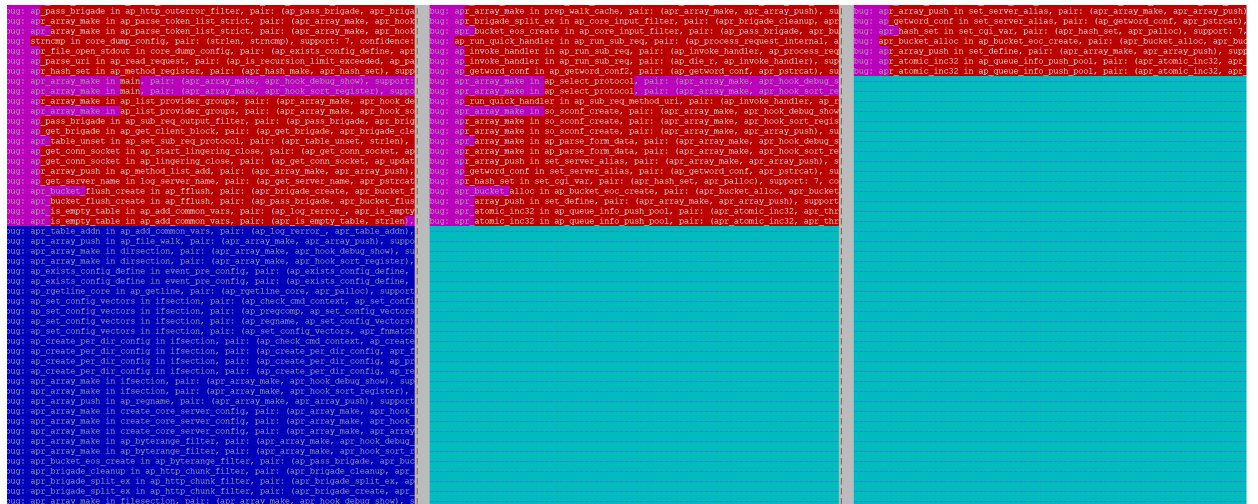


Figure 6: Support = 3, Confidence = 65, Level = 5, ratio = 40

In Figure 6, the bugs detected by each algorithm are 253, 180, and 165 respectively.

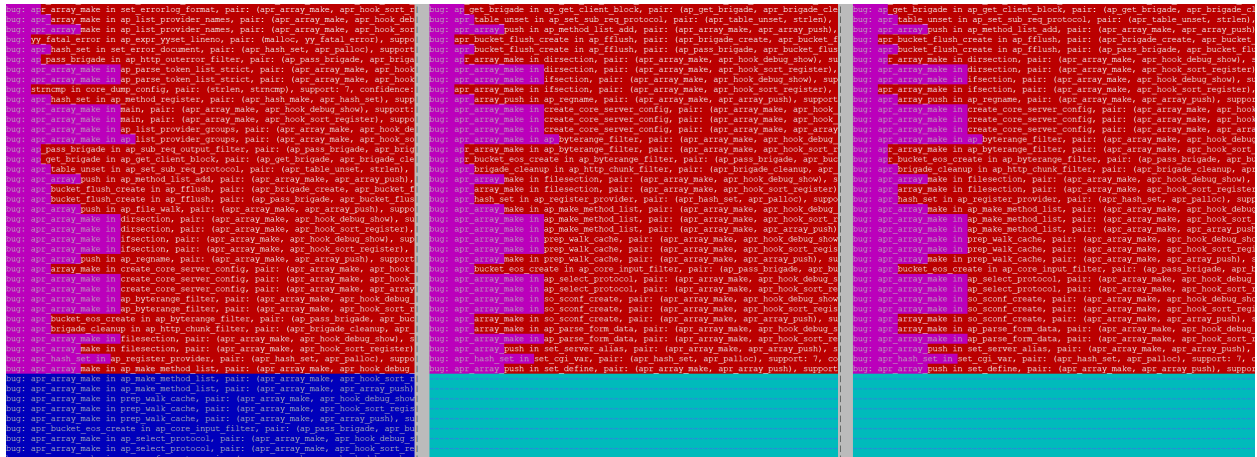


Figure 7: Support = 5, Confidence = 70, Level = 2, ratio = 40

In Figure 7, the bugs detected by each algorithm are 118, 101, and 101 respectively.

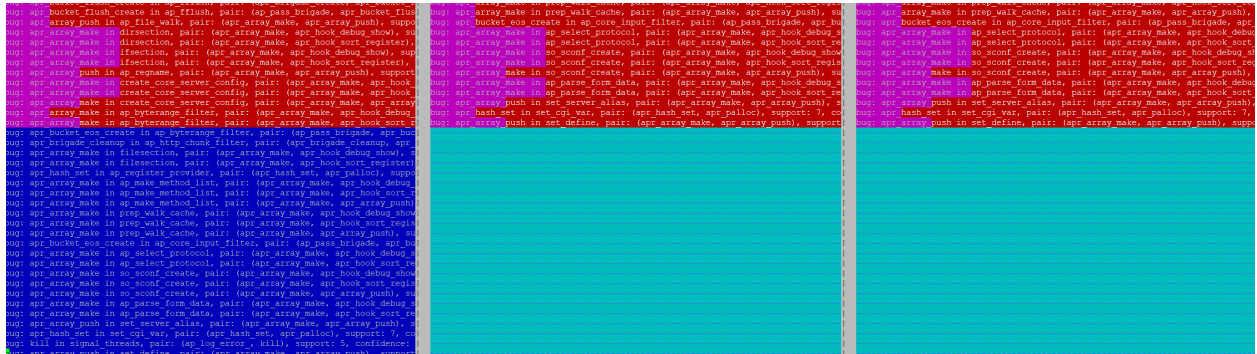


Figure 8: Support = 5, Confidence = 70, Level = 5, ratio = 40

In Figure 8, the bugs detected by each algorithm are 118, 95, and 95 respectively.

In conclusion, our algorithm, *ELISP*, eliminates some bugs that are highly possible to be false positives in cases with lower support and confidence settings. In addition, with a higher unpaired partner ratio, more false positives are supposed to be eliminated.

Part 2 (a). We provide our classification and explanation toward 18 warnings here.

1. CID: 10026 Dereference before null check

Triage: Bug

Location: JspDocumentParser.java (line: 1355)

```
int len = attrs.getLength();
```

Explanation:

Since the developer does null check on *attrs* before dereferencing it in function *startElement*,

I believe that the developer forgets to do null check in function *checkPrefixes*, meaning that it is an unintentional bad practice in codes. The fix here is simple that we just need to do null check on *attrs* in function *checkPrefixes* before dereferencing it.

2. **CID:** 10102 Dereference null return value

Triage: Intentional

Location: SSIProcessor.java (line: 304)

Explanation:

The developer intentionally initiates a string with null value and rewrite it when *firstLetter* does not equal -1. There might be a problem because the program may go to false branch, resulting in returning null pointer. To remove the warning, the developer should use other characters to initialize a string or do null check before returning.

3. **CID:** 10148 Resource leak

Triage: Bug

Location: JkMain.java (line: 455)

Explanation:

The developer does not close or reclaim a system resource after creating a *FileInputStream*, which causes to memory leak. A fix here is to store the new stream with an object name. After loading it in try statement, close it with try-catch in finally block.

4. **CID:** 10167 Dereference before null check

Triage: False Positive

Location: DeltaRequest.java (line: 199)

Explanation:

The warning is not really pointing a bug since the null check is just for checking whether the initialization of *action* in line 65 is successful. There is no line fundamentally rewriting *action* pointer before dereferencing it but adding or removing elements in the list. Therefore, the warning is no need.

5. **CID:** 10176 Dereference null return value

Triage: Bug

Location: SimpleTcpReplicationManager.java (line: 416)

Explanation:

Function *readSession* returns null only when an exception is caught within. However, if an exception happens in Function *readSession*, there is no null check after the function returns, causing the program to crash due to dereferencing *session* which is null. A fix here is doing null check before dereferencing it.

6. **CID:** 10231 Resource leak

Triage: Bug

Location: ChannelNioSocket.java (line: 413)

Explanation:

The developer forgets to close *ssc* before returning at line 413. A way to fix it is closing it before returning.

7. **CID:** 10291 Dereference null return value

Triage: Intentional

Location: SSIServletExternalResolver.java (line: 358)

Explanation:

The developer intentionally returns a null pointer and pass it as an argument to method *startsWith*, which highly make the program to crash. To fix it, the developer should modify function *getContextPath* so it will not always return a null pointer.

8. **CID:** 10351 Dereference after null check

Triage: False Positive

Location: SimpleTcpCluster.java (line: 905)

Explanation:

By observation, I am sure that the developer wants to exclude the case when *message* is null. Although I do not really know the return value of *log.isDebugEnabled()* when object *message* is null, I believe the the developer clearly know it when checking it before dereferencing object *message*. Thus, I consider the warning as false positive.

9. **CID:** 10381 Dereference null return value

Triage: Bug

Location: VirtualDirContext.java (line: 193)

Explanation:

There is no null check before dereference on *files* which might be null in line 192. To fix it, do null check before between line 192 and 193.

10. **CID:** 10395 Unguarded read

Triage: False Positive

Location: StandardWrapper.java (line: 667)

Explanation:

By observation, the function *getServlet* which accesses shared value *instance* is just returning it for function *backgroundProcess* to check the current value of it, which has no intention to compete with other threads. Therefore, I think there is no need for this warning.

11. **CID:** 10396 Dereference after null check

Triage: False Positive

Location: FarmWarDeployer.java (line: 153)

Explanation:

If *econtainer* is null, then the if statement in line 151 will absolutely go to true branch, meaning the program will return instead of dereferencing an object assigned from it. Thus, the warning is no need.

12. **CID:** 10435 Resource leak

Triage: Bug

Location: JDTCompiler.java (line: 235)

Explanation:

The function *isPackage* returns True or False without closing Inputstream *is*, which causes memory leak. To fix it, do null check and store the result in a variable, then close the stream and return the variable.

13. **CID:** 10453 Dereference after null check

Triage: Bug

Location: ChannelUn.java (line: 116)

Explanation:

The warning is reasonable since the developer actually does null check before using super with the method, which implies *wEnv* might be null. The developer does nothing to handle a null condition for *wEnv* and just dereferences it in a parent class method, which is dangerous. To remove the risk, the developer should do something to *wEnv* after the null check before dereferencing it in parent class methods.

14. **CID:** 10507 Dereference after null check

Triage: Intentional

Location: MessageBytes.java (line: 330)

Explanation:

By observation, there are several checks in the following format

```
if ( strValue==null && s!= null )
```

showing that the developer just wants to handle this case while ignoring other cases. This is very dangerous for the latter codes because *strValue* is dereferenced and *s* is passed as an argument in next line. Thus, I think it is a bad practice and should be replaced by

```
if ( strValue==null || s==null )
```

to avoid any possible risks.

15. **CID:** 10514 Missing call to superclass

Triage: False Positive

Location: ChannelCoordinator.java (line: 88)

Explanation:

There is no need to call superclass if we do not need any actions implemented in superclass method. So, I think the program logic is up to the developer and there is no necessity for this warning.

16. **CID:** 10625 Dereference before null check

Triage: False Positive

Location: FarmWarDeployer.java (line: 159)

Explanation:

This is the extension of warning 10396. Based on previous analysis, we know that it is not possible to dereference *engine* if it is a null pointer after the check in line 151. Therefore, the warning can be ignored.

17. **CID:** 10654 Dereference after null check

Triage: False Positive

Location: JDBCStore.java (line: 927)

Explanation:

With the if statement in line 669, *preparedRemoveSql* must not be null, meaning it is not possible to dereference *preparedRemoveSql* in function *remove* and encounter *NullPointerException*. Hence the warning is false positive.

18. **CID:** 10689 Unguarded read

Triage: False Positive

Location: DeltaSession.java (line: 159)

Explanation:

In occurrences block, there are five example pairs of lock and access. In those five pairs, there are two cases, B and C, accessing *deltaRequest* for null check. One is reported as unguarded read while the other is not, which is nonsense. If accessing *deltaRequest* for null check needs to be locked like other three cases, then there should be two warnings here. Thus, I consider this warning as false positive.

Part 2 (b). There are only two defects detected by Coverity and the causes are the same – retrieving all keys and then accessing the map again to get the value for some of the keys. The warning here is indicating that it is more efficient to use an iterator on the entrySet of the map, avoiding the *Map.get(key)* lookup. Take CID 10772 for example. Instead of writing codes below

```
for (String caller : callSites.keySet()) {  
    Set<String> calleeSet = callSites.get(caller);
```

we can do the same thing in another way like below

```
for (Map.Entry<String, Set<String>> entry : callSites.entrySet()) {  
    Set<String> calleeSet = entry.getValue();
```

Although the program without the defects may probably run faster, there is no considerable impact that we need to care about in the program. Therefore, those defects can be classified as false positive.