Design of the TOIF Adaptors and Framework.

Table of Contents

1 Summary

This document is the design of the TOIF adaptors and framework.

2 Introduction

There are two main roles of the TOIF (Tool Output Integration Format):

- Convert the different reports from 3rd party vulnerability detection and scan tools to a common report format, without modifying the original vulnerability detection tool in any way.
- Extract key data from the outputs and supplement the findings with additional weakness information.

The role of the TOIF (Tool Output and Integration Framework) adaptors are to map the findings of 3rd party vulnerability detection and scan tools, to a normalized TOIF output as defined in the specification of the TOIF XML schema. These adaptors will attempt to gather key facts from the 3rd party tool's output such as the data element relating to the finding, the location in the code where the finding was found, and the weakness description. In addition to gathering facts from the tool output, the adaptors will map the weakness to a Common Weakness Enumeration (cwe.mitre.org) for further, more detailed, information. The adaptors are run by the TOIF adaptor framework which provides utilities for constructing the findings and produces the final output. By using a common framework, the Adaptors can be reduced to only parsing their particular tool's output.

It is essential that the adaptors do not modify the original vulnerability detection tools. The internal working of the detection tools must be preserved, and only the output used. To do this the adaptors will wrap the scan tools, initiating their process, and collecting their results. The adaptors will then construct their own output before writing this to disk.

3 Context

This section describes the context in which the adaptors and their framework operate. This is to provide a brief overview before discussing the inputs and outputs of the TOIF adaptors.

The aim of the adaptor, when run on a particular source file, is to capture the output of a particular vulnerability tool and translate its output into a common output format.

The adaptors are to be run during the build of the project in question. This is done to allow access to all the files used during the build and also any required arguments, such as include paths, which the adaptors may need. To achieve this in an unobtrusive manner, a wrapper should be constructed around the compiler. In this way, arguments originally intended for the compiler can be redirected via the adaptor process before being forwarded onto the compiler. By wrapping the compiler, instead of being inserted into the makefile, the requirement of not intruding into the original source or build can be maintained.

The TOIF XML outputs produced from each input source files are written to subdirectories relating to which Adaptor produced them. These files can then be used to integrate the TOIF information into KDM.

4 Configuration

In order to run the adaptors, there is some configuration that will be required. It is essential that the required vulnerability detection tools are pre-installed and have been added to path.

House Keeping:
Because the TOIF output provides information relating to the parties and tools involved in the analysis of a project, this information must be entered into a file and handed to the Adaptor as an argument. The text file is a map of keys and values. The keys relate to facts and entities used in the project, while the values are information about that fact or entity.

House Keeping Example:

```
############################
#         Facts
############################


TOIFSegmentIsRelatedToProject=project1
TOIFSegmentIsProducedByOrganization=org1
TOIFSegmentIsOwnedByOrganization=org1
```

```
TOIFSegmentIsGeneratedByPerson=person1
TOIFSegmentIsSupervisedByPerson=person1

PersonIsInvolvedInProjectAsRole=person1;project1;role1
OrganizationIsInvolvedInProjectAsRole=org1;project1;role2
OrganizationIsPartOfOrganizationAsRole=org1;org2;role2
PersonIsEmployedByOrganizationAsRole=person1;org1;role1

############################
#        Entities
############################

SegmentDescription=Segment relating to the findings on the wireshark project.

#projectId=name;description
project1=Wireshark;Using Wireshark to test the TOIF tool chain.

#personId=name;email;phone
person1=Adam Nunn;adam@kdmanalytics.com;555-1234
#person2=Joe Bloggs;blogs@kdmanalytics.com;555-1234

#organizationId=name;description;address;phone;email
org1=KDM;Kdm Analytics;Richmond Road;555-5555;kdm@kdmanalytics.com

#organizationId=name;description;address;phone;email
org2=Acme Corporation;Acme Corporation;blah;555-5555;thingamy@thingamy.com

#roleId=name;description
role1=Software Developer;Developer on the TOIF Adaptors project
role2=company;employer
```

Although most entities and facts within the TOIF XML schema are optional, there are few house keeping elements which are considered vital.

- TOIFSegmentIsRelatedToProject
- Project
- TOIFSegmentIsGeneratedByPerson
- Person

Running the Adaptor from command line:

The adaptor is run from the command line. Which adaptor to use with the framework is included in the arguments.

```
	java –cp "<classpath>" ToifAdaptor --adaptor <Adaptor Name> --inputFile <full path to input file> --outputDirectory <path to output directory> --houseKeeping <path to house-keeping file> [Additional arguments]
```

“-cp”

The class path must be provided in order for the framework to find the correct adaptor to run. This lets the Java Virtual Machine know where to look for the classes and packages. Multiple paths can be provided by separating the paths with a colon “:”. Note: On windows, if globbing is used, you may need a semi-colon after the classpath. Eg: java "C:\Users\user\adaptors*";

“--adaptor” or “-a”

The name of the adaptor class. This is the adaptor that is to be used with the input source file. From this class, the framework is able to discover house keeping facts about the adaptor as well as which generator to call and what options to use. Typically the name is without the adaptor version number.

“--inputFile” or “-i”

The full path to the input source file. In order for the adaptors to create all the facts for this file, a full path must be provided

“--outputDirectory” or “-o”

The path to the output directory. This is the directory where the subdirectories containing the TOIF XML file will be written.

“--houseKeeping” or “-h”

The path to the file containing the facts about the project's house keeping. This file is specific to

each adaptor and each project. This is because it is down to the user to provide the project details as well as which generator (scan tool) is running on the system.

Any additional arguments may be entered after the ToifAdaptor's required arguments. These may be included files or compilation options, they will vary from generator to generator.

eg splint takes -I and -D options: java -cp "/home/user/adaptors/*" ToifAdaptor -a SplintAdaptor -i /home/user/foo.c -o /home/user/output -h housekeepingFile.txt -I./includes -D_U_=

Integrating with C project's build:
The best way to integrate the adaptors into the build, is by wrapping the compiler and the adaptors into a script. When the compiler is called, the adaptors will be run for every source file used. To get the build process to use this wrapper instead of the compiler on its own, the compiler flag needs to be set during configuration of the make:

```
./configure CC=myGccWrapper
```

The make can then be continued as normal:

```
make
make install
```

Integrating with Java project's build:
It may be possible to integrate into a Java project's build by modifying Apache Ant's “build.xml” file. Create a new target which will find all the “.class” files in the destination directory of the project. For each file, the javaAdaptors script will be run with the arguments that are specified:

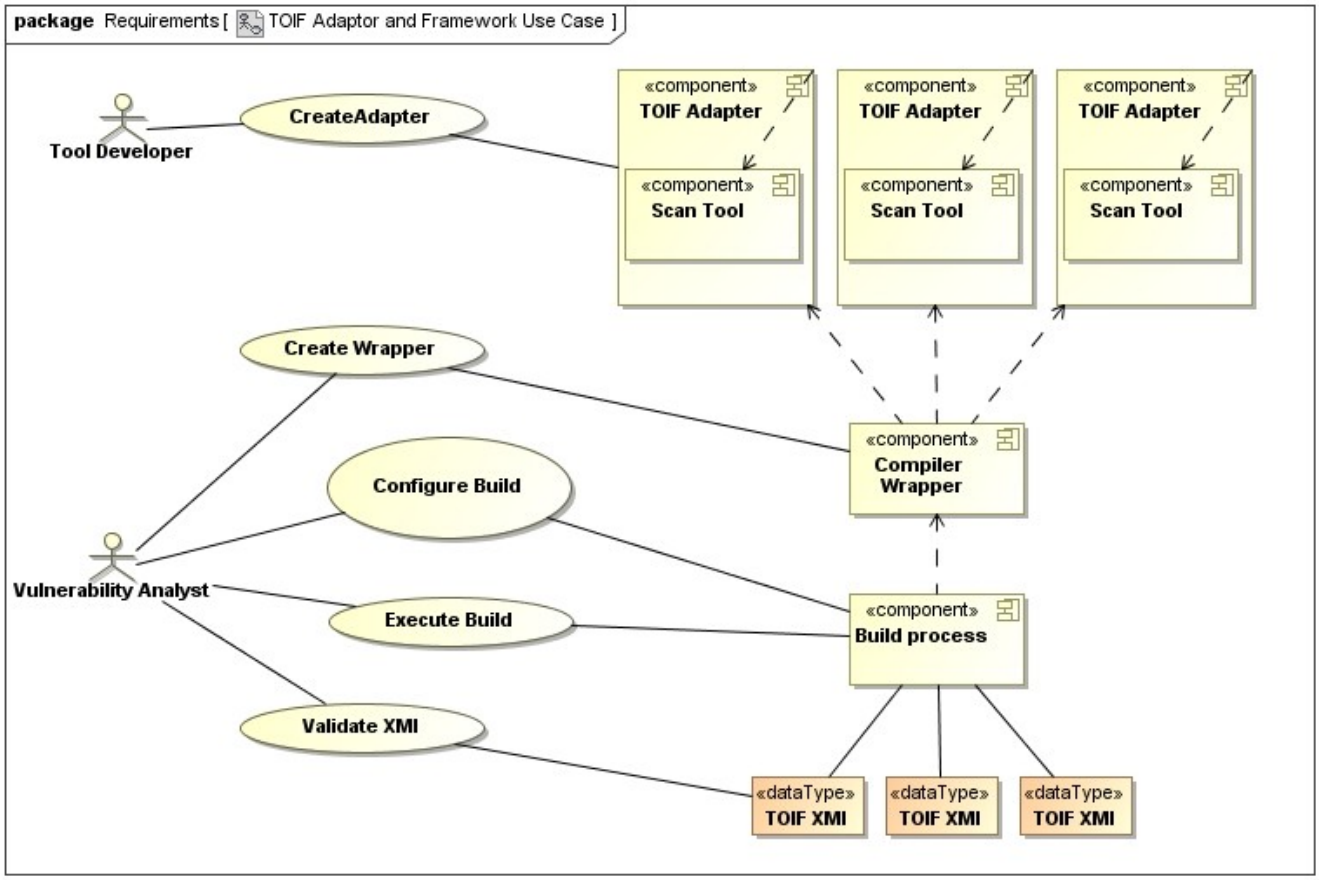
```
<target name="check" depends="compile_src">
    <foreach param="file" target="run">
            <path>
                    <fileset dir="${classes_dir}">
                        <filename name="**/*.class" />
                    </fileset>
            </path>
        </foreach>
</target>
```

```
<target name="run" >
    <exec executable="python">
            <arg value="C:/Users/user/javaAdaptors.py"/>
        <arg value="${file}"/>
                <arg value="${build_dir}"/>
        </exec>
</target>
```

To run the adaptors and compile the project, the following command should be given:

```
    ant check
```

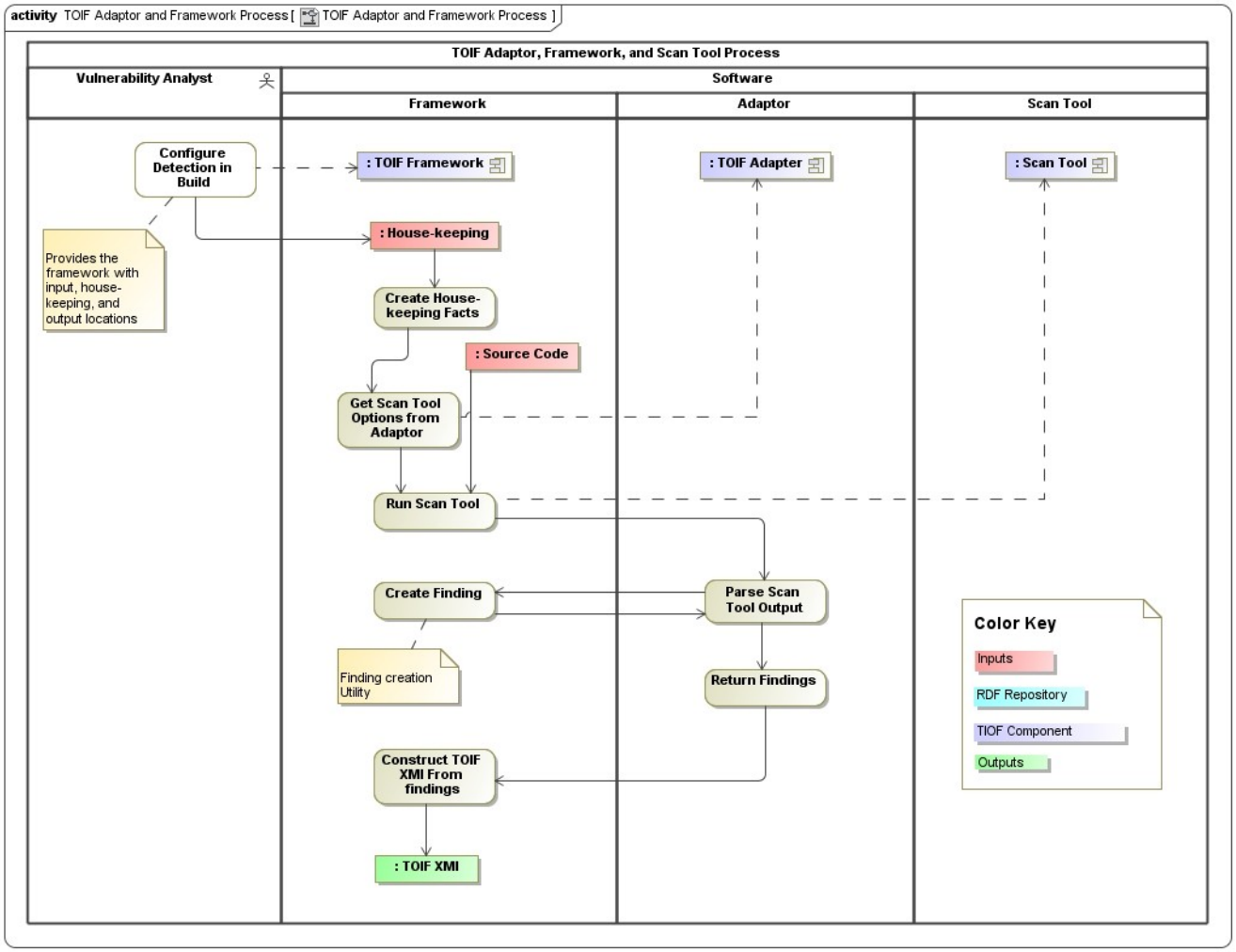
5 Use Case



- Tool integrator creates an Adaptor for each Scan Tool.
- The vulnerability analyst creates a wrapper for the compilers which will be used during the build.
- The build is configured to use the wrapper instead of the original compiler.

- The build is run, producing the TOIF output.
- The TOIF output is validated against the TOIF XML schema.

6 Adaptor and Framework Interaction



Framework:
The framework is run with the adaptor as an argument. From this argument, the Framework can find the adaptor class and query that class to find information out about the scan tool, the adaptor's name and description, and the arguments with which to run the vulnerability detection tool.

With the house keeping file as an argument the framework can create the elements which describe the project in detail.

The framework initiates the scan tool with the proper arguments, provided by the adaptor and the user, and hands this running process to the adaptor for parsing.

Once the adaptor has returned the findings it created whilst parsing the scan tool's output, the framework can create the XML and write it to disk.

Adaptor:
The adaptor takes the running process from the framework and parses the output of the scan tool. Because each vulnerability detection tool produces it own unique output, each adaptor must be different in order to handle each individual output format.

When the adaptor discovers a finding, it passes the key elements of the finding to the finding creator utility in the framework package. This finding creator utility returns the facts and elements which make up the finding.

Finally the adaptor returns all the finding elements it has found to the framework.

7 Creating an Adaptor

All adaptors must extend the abstract class “AbstractAdaptor”, in the framework package. This class contains a series of methods which must be implemented. The majority of the methods return strings describing various elements in the TOIF. There are however a couple of more involved methods:

public abstract ArrayList<Element> parse(Process process, AdaptorOptions options, File file);
The returned ArrayList is a list of all the elements created by passing information extracted from the process to the finding creator utility. How the Adaptor parses the output of the scan tool (process) is unique to each particular scan tool.

public abstract String[] runToolCommands(AdaptorOptions options, String[] otherOpts);
This method must return a String array of the arguments required to run the generator tool. If the tools output must be in a particular format, now is the time to specify it. The String array “otherOpts” should be merged into the commands if other ability to pass other options such as, include paths, is required.

Example: {"rats", "--xml", "--quiet", options.getInputFile().toString()}

In addition to completing the abstract class, a configuration file must be created in a folder called “config” in the root of the jar. The name of the configuration file must be of the form:

```
/config/<Adaptor class simple name>Configuration
```

Within this file there must be a map of weakness Ids, their Clusters, CWEs, and SFPS. It was decided that the CWEs and SFPs must be defined together, however, the SFPs are mapped to the clusters from within the framework.

```
#possible dereferencce of null pointer
nullderef=Memory Access;SFP-7;CWE-476
```

8 Tool Output

The output of the adaptors will conform to the TOIF XML schema. This schema describes the structure of the elements within the TOIF XML file.

For each individual input source file, an output TOIF XML file is created. This is written to a directory relating to the adaptor which generated it.

```
.../<output directory>/<adaptor name>/<input filename>.toif
```

To make the files more concise, most of the house keeping elements are extracted into a separate common file, “GENERAL_INFORMATION.toif”. One of these files is created for each adaptor.

This output can then be used by other tools to generate reports or integrated into KDM.