

UNIVERSITY OF SANTIAGO DE  
COMPOSTELA



ESCOLA TÉCNICA SUPERIOR DE ENXEÑARÍA

## Improvements in IDS: adding functionality to Wazuh

*Autor:*

**Andrés Santiago Gómez Vidal**

*Directores:*

**Purificación Cariñena Amigo  
Andrés Tarascó Acuña**

**Computer Engineering Degree**

**February 2019**

Final degree project presented at the Escola Técnica Superior de Enxeñaría of  
the University of Santiago de Compostela to obtain the Degree in Computer  
Engineering





**Ms. Purificación Cariñena Amigo**, Professor Computing Science and Artificial Intelligence at the University of Santiago de Compostela and **Mr. Andrés Tarascó Acuña**, Managing Director at Tarlogic Security S.L.

STATE:

That the present report entitled *Improvements in IDS: adding functionality to Wazuh* written by **Andrés Santiago Gómez Vidal** in order to obtain the ECTS corresponding to the final degree project of the Computer Engineering degree was conducted under our direction in the department of Computer Science and Artificial Intelligence of the University of Santiago de Compostela.

For the purpose to be duly recorded, this document was signed in Santiago de Compostela on February TODO, 2019:

The director,

The codirector,

The student,

(Purificación Cariñena Amigo) (Andrés Tarascó Acuña) (Andrés Santiago Gómez Vidal)



# Index

<b>1</b>	<b>Requirements</b>	<b>1</b>
1.1	Use cases . . . . .	1
1.1.1	Use cases actors . . . . .	2
1.1.2	Use cases list . . . . .	2
1.2	Requirements analysis . . . . .	2
1.2.1	Non functional requirements . . . . .	2
1.2.2	Functional requirements . . . . .	2
1.2.3	Domain requirements . . . . .	2
<b>2</b>	<b>Planning</b>	<b>3</b>
2.1	Initial WBS . . . . .	3
2.1.1	WBS dictionary . . . . .	4
2.2	Initial planning . . . . .	6
2.3	Final planning . . . . .	16
2.4	Risk management . . . . .	17
2.4.1	Risk metrics . . . . .	17
2.4.2	Risk types . . . . .	18
2.4.3	Risk identification . . . . .	18
2.4.4	Risk analysis and planning . . . . .	19
2.4.5	Risk supervision . . . . .	28



# List of Figures

2.1	Initial planning . . . . .	7
2.2	“Beginning of the project” planning . . . . .	8
2.3	“Increment 1: Common attacks in Windows Server” planning . .	9
2.4	“Increment 2: Use of data from Sysmon” planning . . . . .	10
2.5	“Increment 3: Detection/action against ransomware” planning . .	11
2.6	“Increment 4: Adapt Wazuh configuration to typical requirements from enterprises” planning . . . . .	12
2.7	“Increment 5: Explore solutions in problems with GPDR” planning	13
2.8	“Increment 6: Additional detection for GNU/Linux” planning . .	14
2.9	“Increment 7: VirusTotal integration” planning . . . . .	15
2.10	“Closing of the project” planning . . . . .	16





# List of Tables

2.1	Probability classification of risks . . . . .	17
2.2	Impact classification of risks . . . . .	17
2.3	Method of calculation of exposition based on probability and impact	17
2.4	List of the risks of the project . . . . .	18



# Chapter 1

## Requirements

The requirement specification is a full description of the software the project is to develop.

PMBOK[1] states that requirements are conditions or capabilities that a product must meet to satisfy the contract. The requirements expose the needs of the client, which have to be accomplished to finish the project successfully. In this project the requirements will be fulfilled in multiple stages along the project.

Note that the client in this case is Tarlogic even if the product is a contribution to an open source project.

This specification contains:

- **Use cases:** Functionalities that the software will provide.
- **Requirements:** Depending of their type they can describe features, data, relations, properties or any details necessary to explain the system without ambiguity, in a way it can be easily understood.

In this project the functional requirements are not included because they can be considered a redundant version of the use cases, because both describe the same functionalities. Uses cases were chosen over functional requirements because they were considered to be easier to understand and have greater detail. If this project had the need of a very complex requirement specification it would be interesting to have both, as each could help to understand the other better, but in this project the specification should be quite simple.

### 1.1 Use cases

A use case is a description of all the ways an end-user wants to “use” a system. These “uses” are like requests of the system, and use cases describe what that

system does in response to such requests. In other words, use cases describe the conversation between a system and its user(s), known as actors. Although the system is usually automated (such as an Order system), use cases also apply to equipment, devices, or business processes.[2]

#### **1.1.1 Use cases actors**

#### **1.1.2 Use cases list**

### **1.2 Requirements analysis**

#### **1.2.1 Non functional requirements**

#### **1.2.2 Functional requirements**

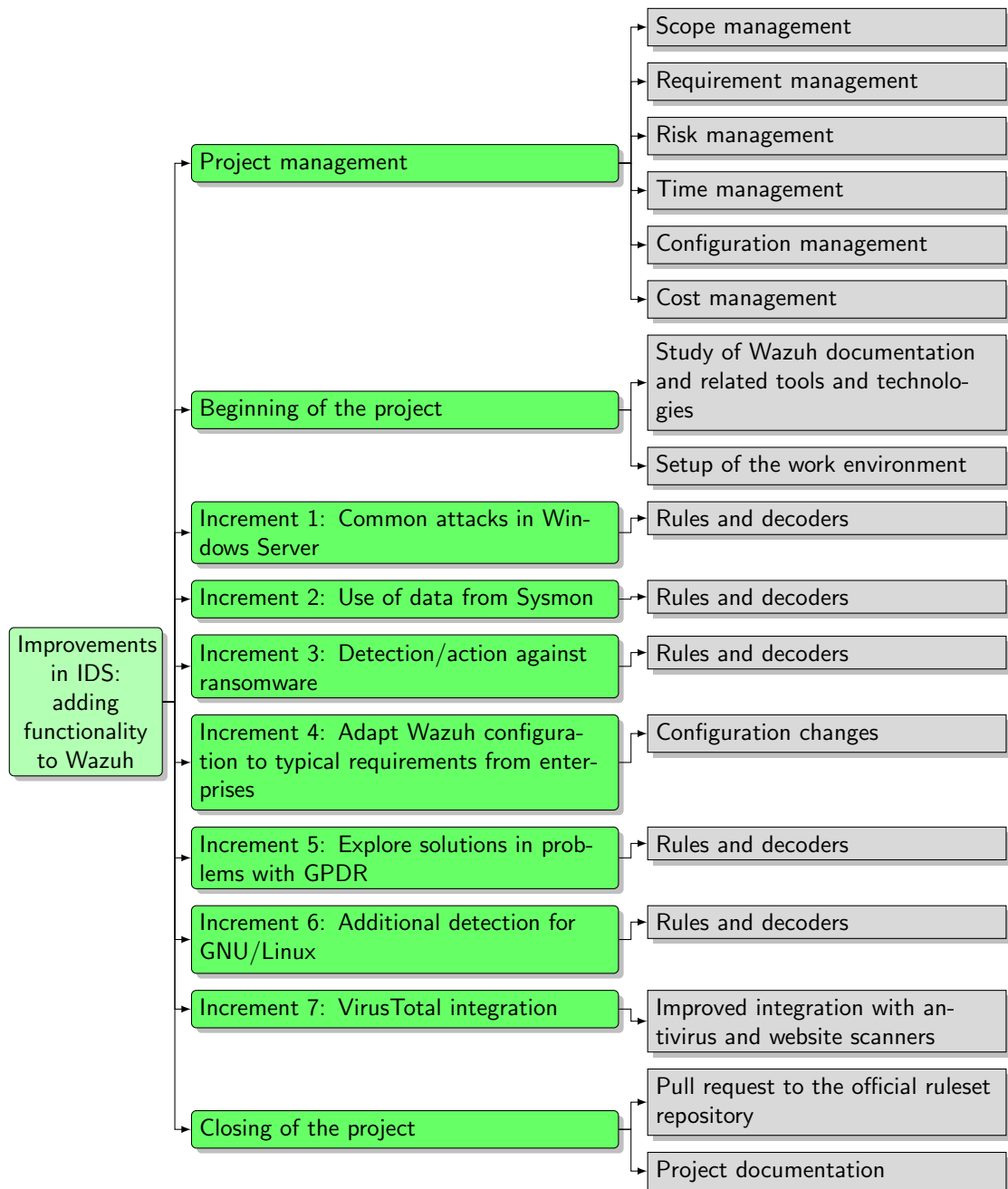
As mentioned before these are omitted because of the redundancy with use cases.

#### **1.2.3 Domain requirements**

# Chapter 2

## Planning

### 2.1 Initial WBS



### 2.1.1 WBS dictionary

#### 1. Project management

- a **Scope management:** Scope explanation, set the restrictions of the project and determine what is going to be turned in at the end of the project.

- b **Requirement management:** Analysis, requirement specification and probably a traceability matrix.
- c **Risk management:** Identification, analysis, classification, planning and supervision of risks.
- d **Time management:** Planning (initial and real), any planning changes and necessary measures.
- e **Configuration management:** Documentation on the management of changes and control version.
- f **Cost management:** Cost estimation (direct and indirect) of software, hardware and resources.

## 2. Beginning of the project

- a **Study of Wazuh documentation and related tools and technologies:** Is the base for multiple aspects of the project and if it is done correctly it can mean less hours in related work.
- b **Setup of the work environment:** Installation and basic configuration of the virtual machines of the project, like having a functional Wazuh environment.

## 3. Increment 1

- a **Rules and decoders:** The objective is to be able to detect common attacks in Windows Server (specifically 2016 and 2019), but it should be backwards compatible and depending on the difficulty it could be worth to ensure support for Windows 10 Pro too. This rules are the final product of this increment, which probably will need more time than any other increment, because its heavy study and testing.

## 4. Increment 2

- a **Rules and decoders:** It will need a preliminary study of Sysmon and the ways to use its data to improve detection in certain situations. It is possible that this increment will modify rules and decoders of the previous one.

## 5. Increment 3

- a **Rules and decoders:** This increment tries to produce rules and decoders to detect ransomware and launch alerts and maybe actions against the attack, like rollback to a previous backup or try to stop the attack from repeating in a short period of time.

## 6. Increment 4

- a **Configuration changes:** Adapt Wazuh to the typical requirements from enterprises. This means that an enterprise could choose from a set of templates, with different security profiles.

## 7. Increment 5

- a **Rules and decoders:** Most should be focused on detecting changes on the protected files. Part of this increment should be the investigation on normal problems of these technologies and recent innovations and solutions.

## 8. Increment 6

- a **Rules and decoders:** There would be preliminary study to do, but the increment should be about expanding the already done work in the field, probably focusing in services and security technologies like SELinux or AppArmor.

## 9. Increment 7

- a **Improved integration with antivirus and website scanners:** The idea is to improve the detection as much as possible with the help of VirusTotal malware scanners, which is updated consistently and so it would mean a consistently updated detection for a system with Wazuh without the need to write new rules and decoders. Obviously there is a difference in the scope and objectives of these technologies, which can be redundant, but this could be certainly interesting in some cases.

## 10. Closing of the project

- a **Pull request to the official ruleset repository:** There is a fundamental need to investigate the correct way to organize the the forked repository for a pull request to an official repository like this. In any case the status of the fork should be checked before and there should be a high amount of commits and use a different branch for each functionality, allowing an easier way to select what to admit or not in the official repository.
- b **Project documentation:** The memory and presentation of the project and whatever other documentation if necessary.

## 2.2 Initial planning

The tasks marked in **red** are essential to the project, meanwhile the ones marked in **cyan** are considered optional and only will be done if there is enough time



left. The tasks marked in **yellow** are normal, and they are used when there is no need to distinguish between essential and optional.

The next Gantt diagram shows the initial planning, from the draft proposal (31/10/2018) to the end of the project (TODO/02/2019).

Furthermore the last two weeks are marked with a grey overlay to mark that there are only about 17 weeks before the due date of this project (in February). This difference is because the estimation of the tasks was made by the student and so it is not reliable, which means that it could be optimistic or pessimist. Thus the need to either reduce tasks or have more that there were expected to fit.

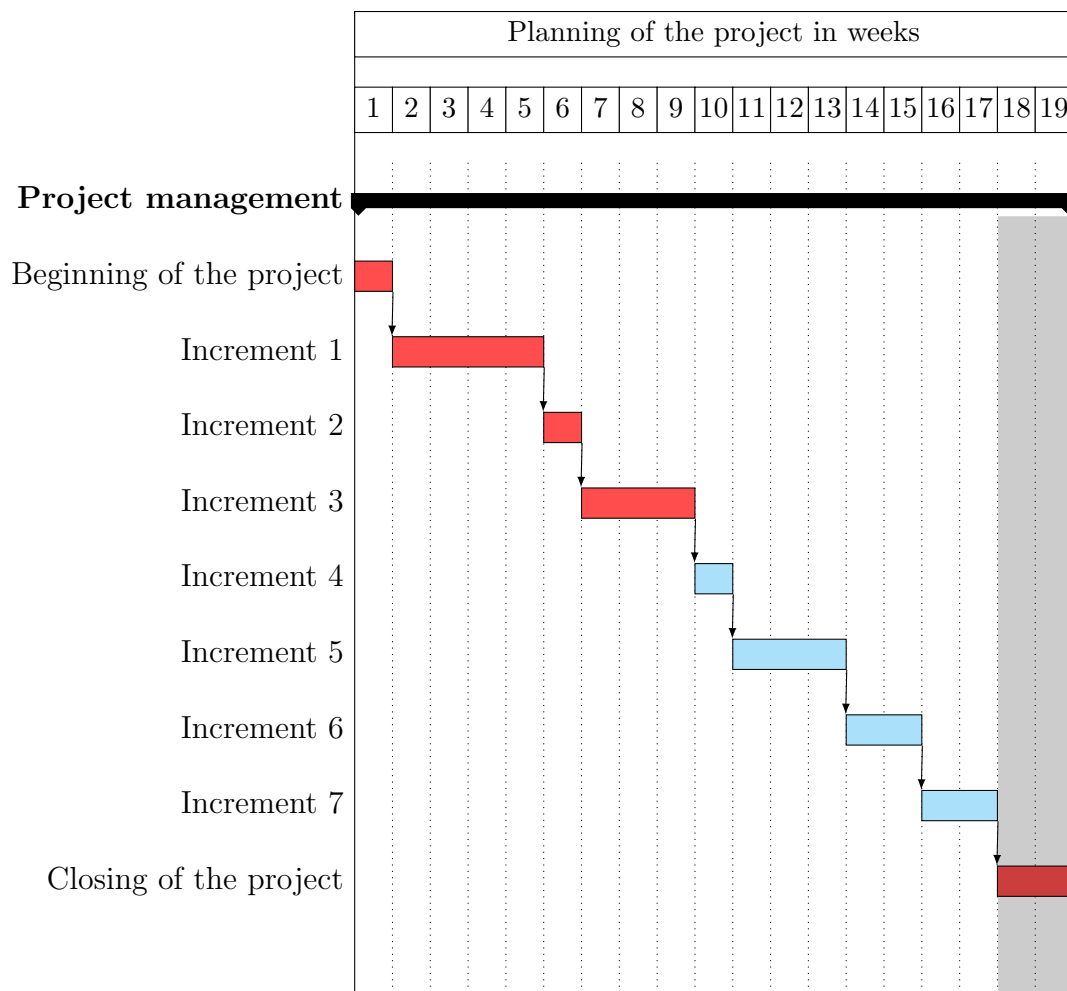


Figure 2.1: Initial planning

The rest of the Gantt diagrams are organized in days, for a more detailed planning.

It is important to note that these plannings could change during the project, either because controlled measures or any unexpected reason.

The order they are implemented could change too and that is the reason because these diagrams have not a set date for start and end, yet.

In other words, they could be described as the models for the final Gantt diagrams.

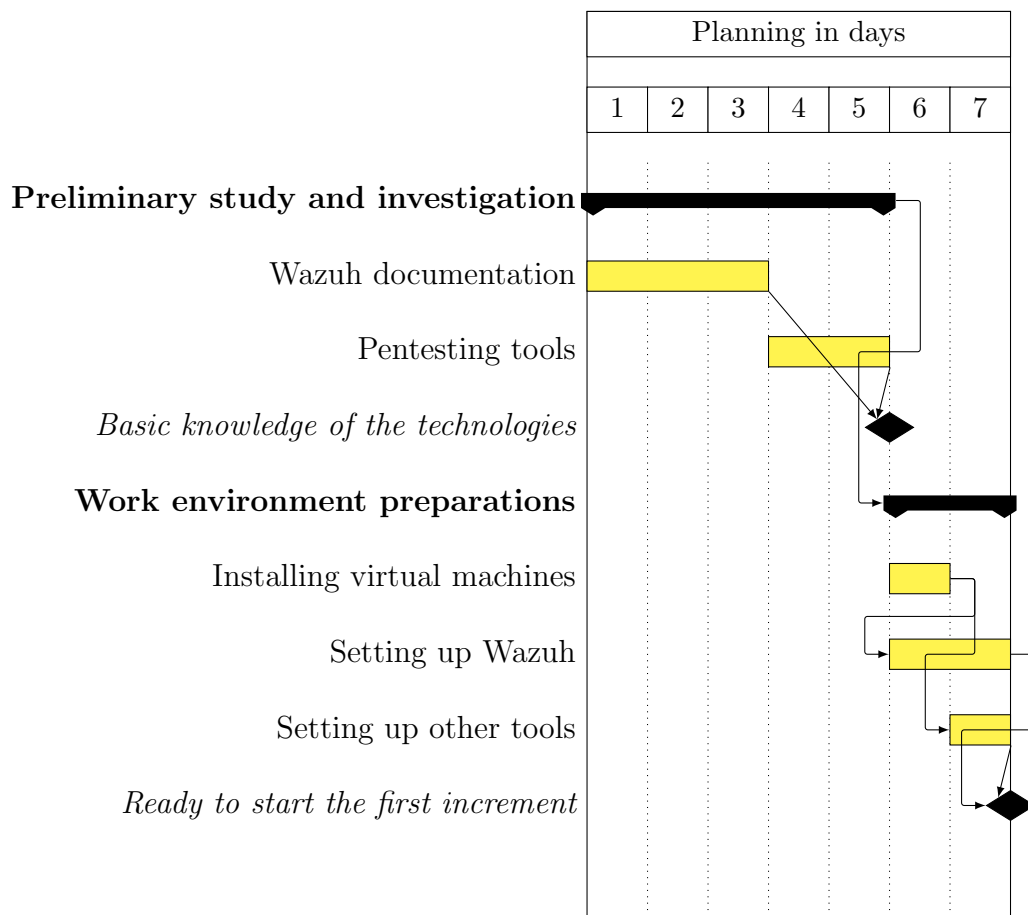


Figure 2.2: “Beginning of the project” planning

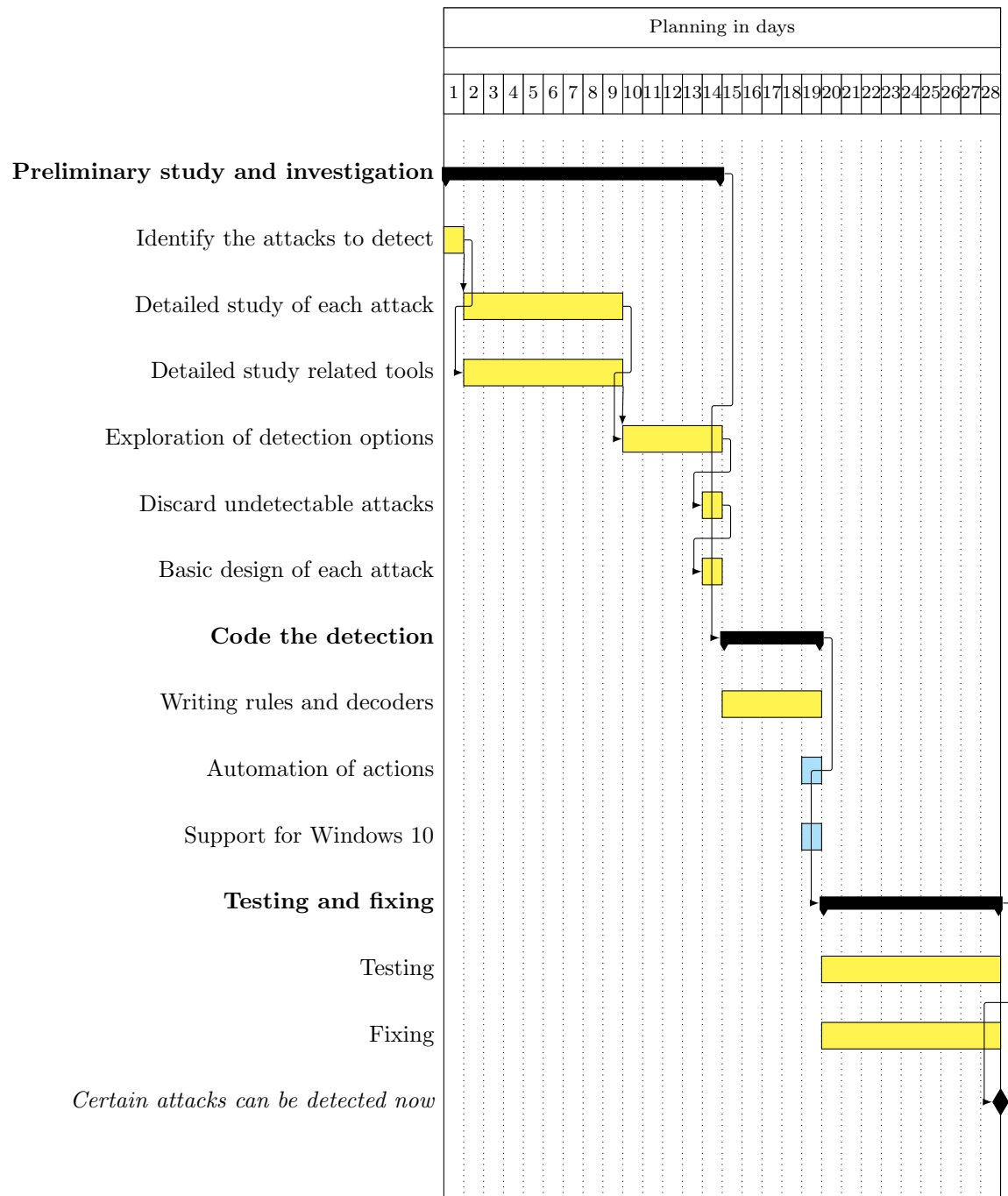


Figure 2.3: “Increment 1: Common attacks in Windows Server” planning

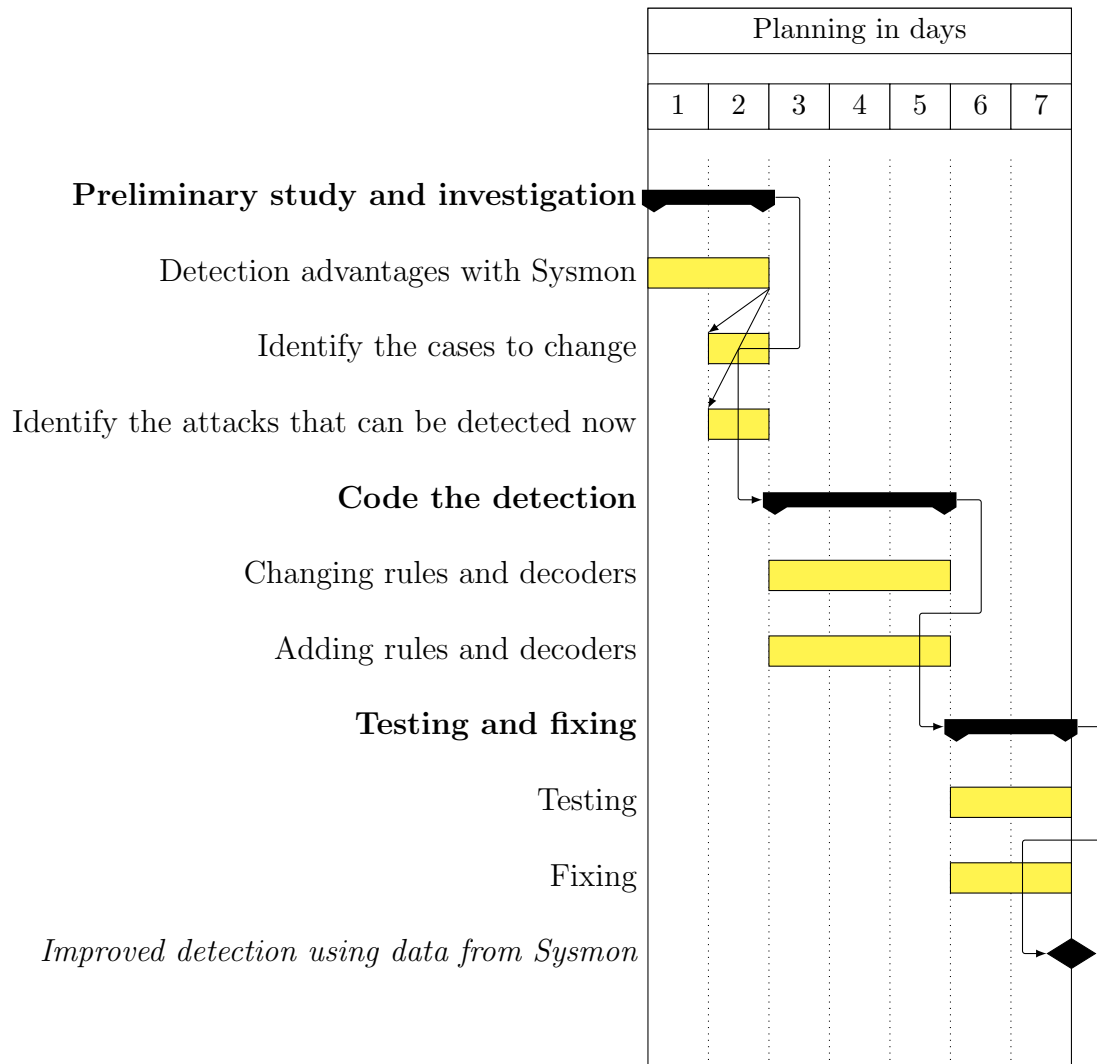


Figure 2.4: “Increment 2: Use of data from Sysmon” planning

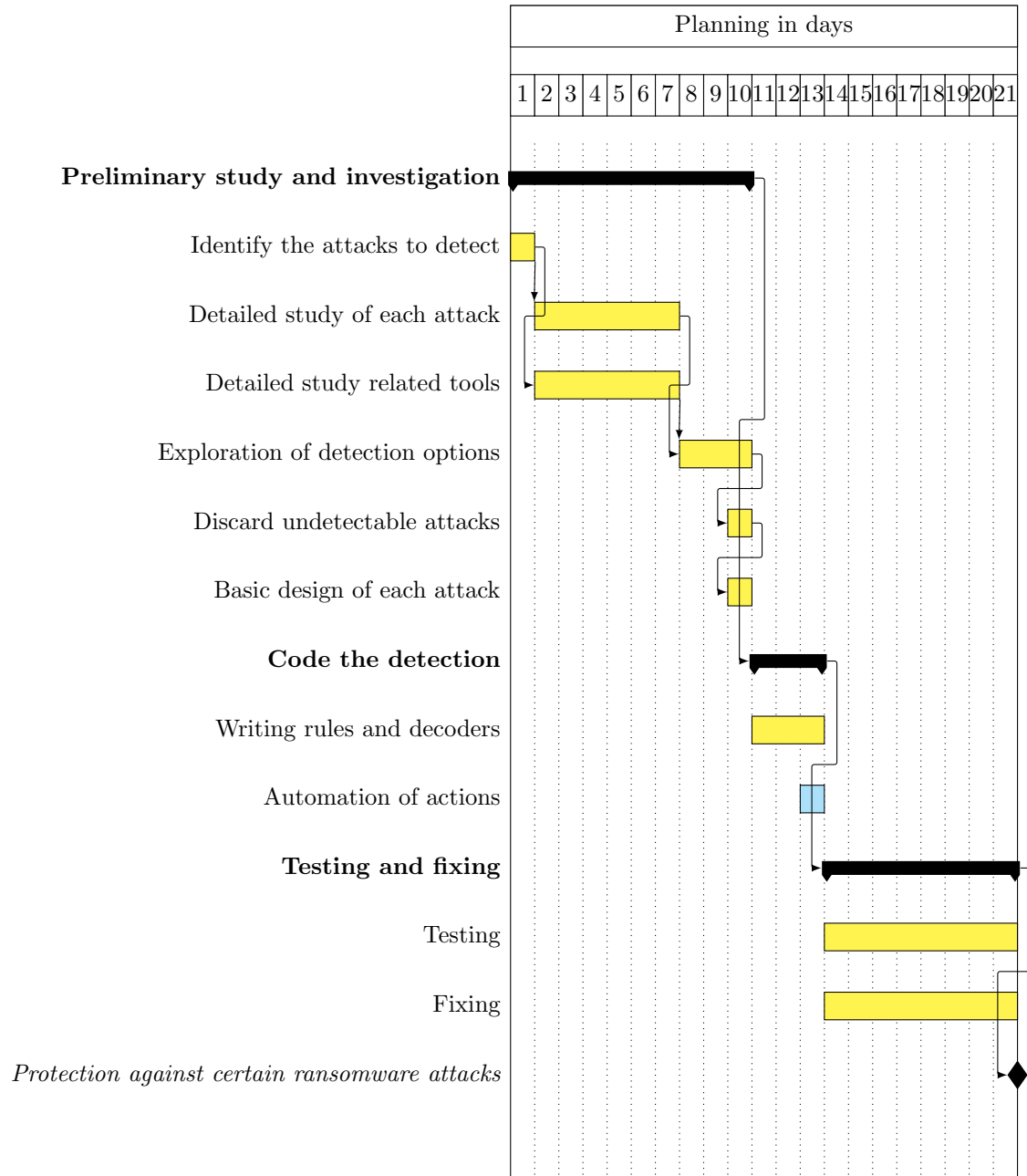


Figure 2.5: “Increment 3: Detection/action against ransomware” planning

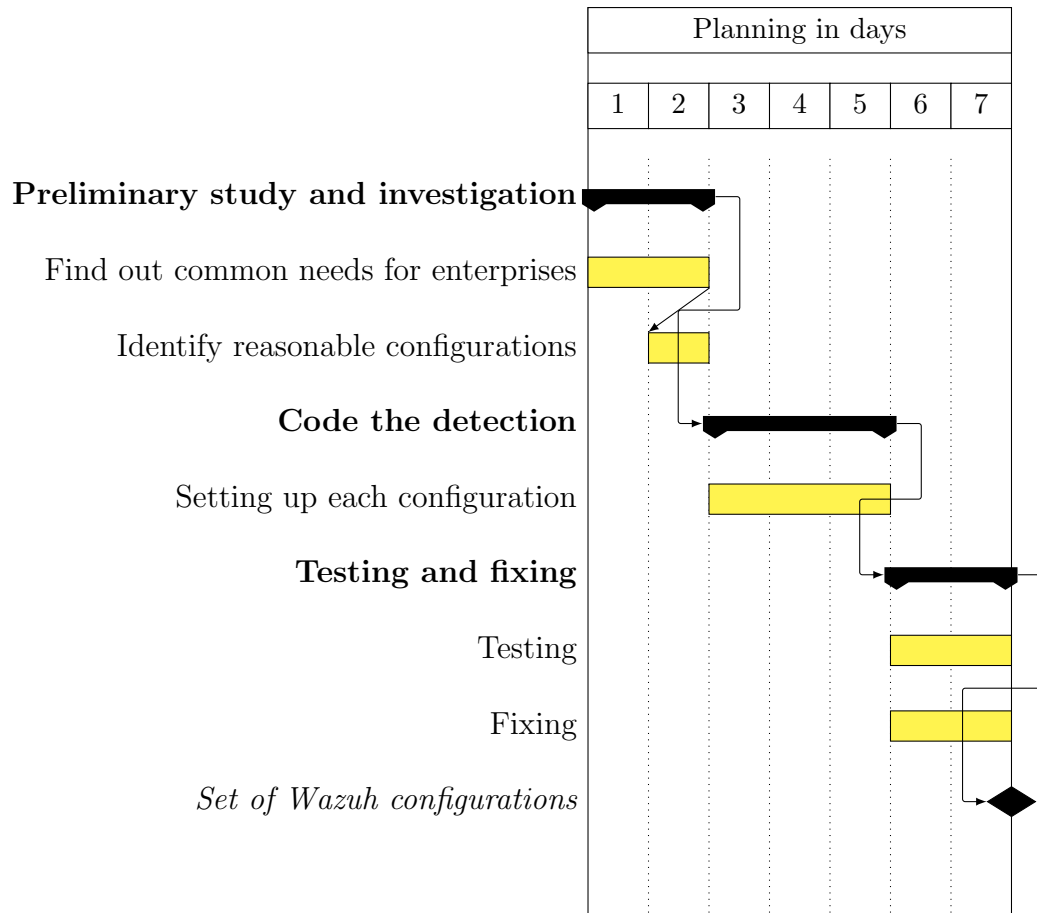


Figure 2.6: “Increment 4: Adapt Wazuh configuration to typical requirements from enterprises” planning



Figure 2.7: “Increment 5: Explore solutions in problems with GPDR” planning

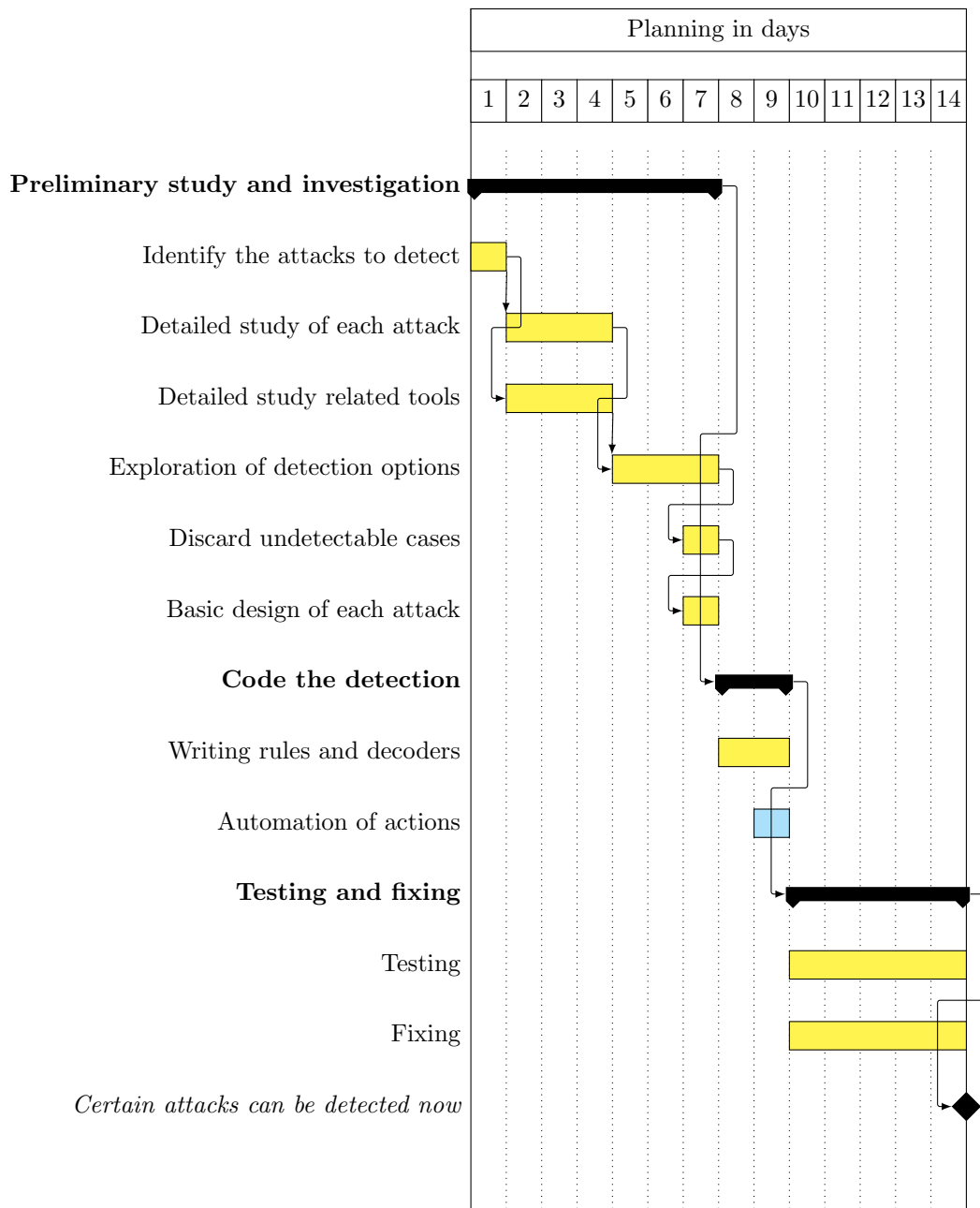


Figure 2.8: “Increment 6: Additional detection for GNU/Linux” planning



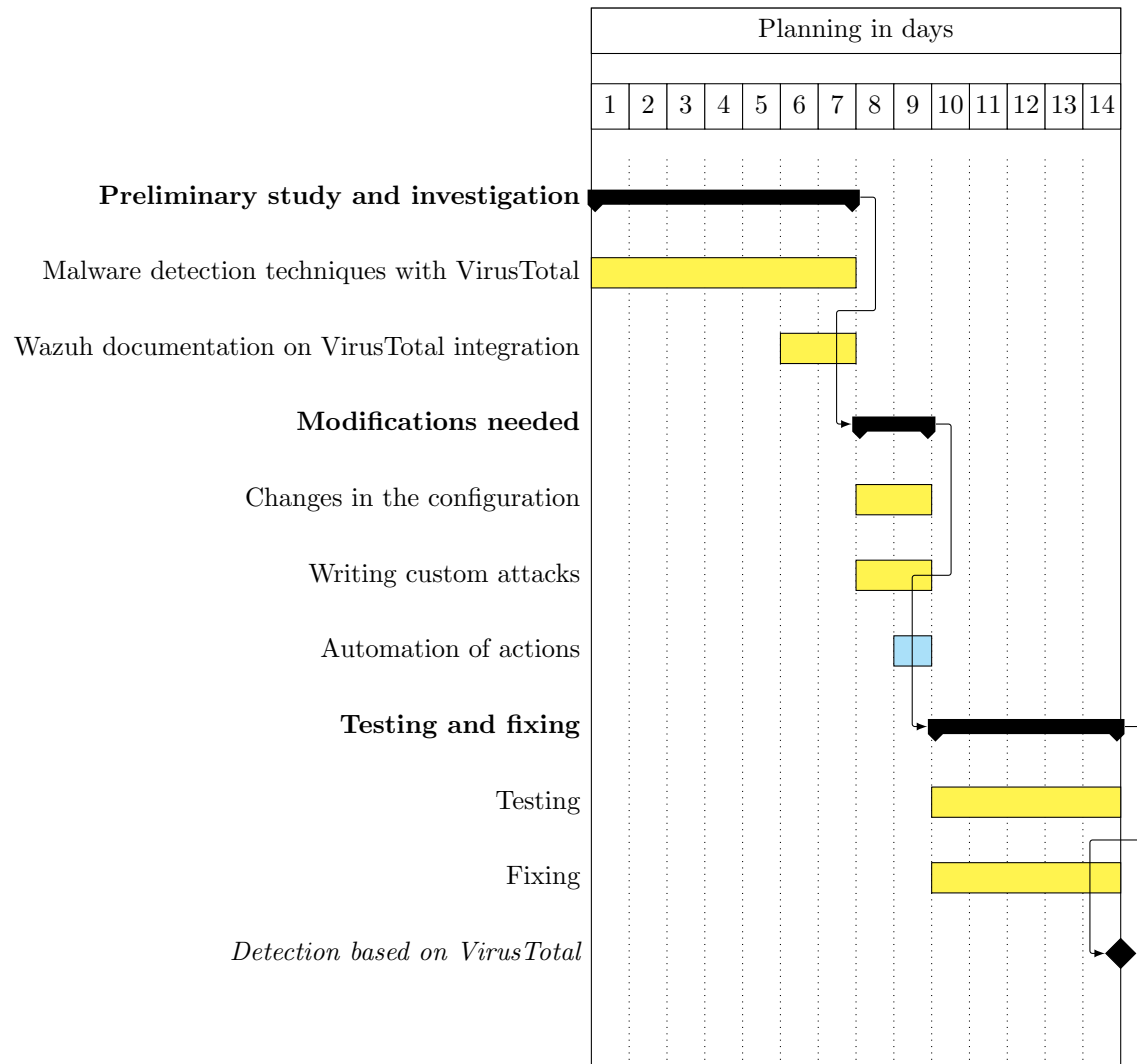


Figure 2.9: “Increment 7: VirusTotal integration” planning

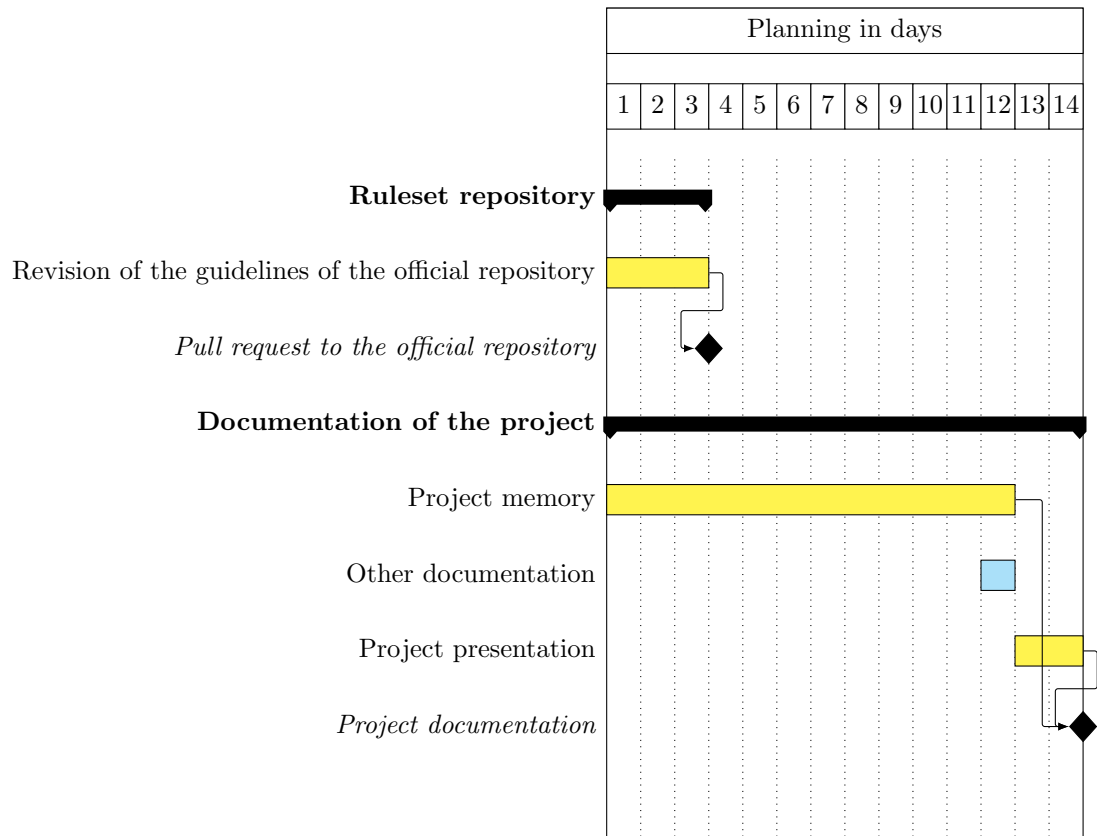


Figure 2.10: “Closing of the project” planning

## 2.3 Final planning

## 2.4 Risk management

### 2.4.1 Risk metrics

Chances of the risk happening	Probability
$\geq 80\%$	High
Between 30% and 80%	Medium
$\leq 30\%$	Low

Table 2.1: Probability classification of risks

Resource in Place / Effort / Cost	Impact
$\geq 20\%$	High
Between 10% and 20%	Medium
$\leq 10\%$	Low

Table 2.2: Impact classification of risks

Exposition		Probability		
		High	Medium	Low
Impact	High	High	High	Medium
	Medium	High	Medium	Low
	Low	Medium	Low	Low

Table 2.3: Method of calculation of exposition based on probability and impact

### 2.4.2 Risk types

### 2.4.3 Risk identification

Identifier	Name
R-01	Optimist planning, “best case” (instead of a realistic “expected case”)
R-02	Bad requirement specification
R-03	Design errors
R-04	Lack of key information from sources
R-05	Lack of feedback or support from the security consultants of Tarlogic
R-06	The learning curve of some technologies is larger than expected
R-07	The unexplained parts of the project take more time than expected
R-08	Can not access source material
R-09	Unexpected changes to any of the software used in the project
R-10	Loss of work
R-11	Wrong management of the project’s configuration
R-12	A delay in one task leads to cascading delays in the dependent tasks
R-13	The student can not find a way to code the detection of a certain occurrence
R-14	The quality of the product is not enough
R-15	Sickness or overwork
R-16	Performance issues
R-17	Unnecessary work
R-18	Optional requirements delay the project

Table 2.4: List of the risks of the project

### 2.4.4 Risk analysis and planning

Identifier	<b>R-001</b>
Name	Optimist planning, “best case” (instead of a realistic “expected case”)
Description	An optimistic planning at the start of the project does not take into account problems or delays, and so it does not allocate time for them.
Negative effects	Could mean the failure of the project if the objectives can not be accomplished in the time left. Cascading delays, because the work done would not fit the planning.
Probability	Medium
Impact	High
Exposition	High
Indicator	There are 3 consecutive delays, after the beginning of the project.
Prevention: Avoid	Allocate a bit more time than initially expected for each task, in case something goes wrong.
Correction: Mitigate	Redo the planning. Reduce the scope of the project, leaving out initially planned optional increments.

Identifier	<b>R-002</b>
Name	Bad requirement specification
Description	The requirements specified at the beginning of the project are not specific enough, are not needed or there are new requirements after the beginning of the project.
Negative effects	Possible failure of the project if the objectives can not be accomplished in the time left. Wasted time, due to lack of communication in the requirement specification.
Probability	High
Impact	High
Exposition	High
Indicator	There are 3 changes in the requirements specification.
Prevention: Mitigate	Confirm that all the requirements have been identified at the beginning of the project. Assure that there is no ambiguity in the requirement specification.
Correction: Mitigate	Redo the requirement specification. Rework of related requirements and work based on them, including the need to test the results. Redo the planning. Reduce the scope of the project.

Identifier	<b>R-003</b>
Name	Design errors
Description	A design is not enough or is incorrect. This can be found in later stages, when it is clear that the implementation based on the design would not satisfy the requirements.
Negative effects	Having to redesign and maybe redo the work based on the design. Minor delays.
Probability	Low
Impact	Medium
Exposition	Low
Indicator	There are 3 designs that need rework.
Prevention: Mitigate	Use design patterns if needed (this project should have very simple designs, so it is possible that there is no need to use them). Make the design as simple and modular as possible.
Correction: Mitigate	Redesign and probably change and test the work based on the design.

Identifier	<b>R-004</b>
Name	Lack of key information from sources
Description	Not having key information from articles, documentation or manuals.
Negative effects	Minor delays. Loss of quality. Added difficulty, even if the work is done in time. Maybe rework and test the functionality, even completely, to follow the desired procedure.
Probability	Medium
Impact	Medium
Exposition	Medium
Indicator	The duration of the study of the attack and the related tools takes 50% than expected.
Correction: Mitigate	Ask the security consultants of Tarlogic for specific information. Possibly the need to rework completely some functionality.

Identifier	<b>R-005</b>
Name	Lack of feedback or support from the security consultants of Tarlogic
Description	Because I do not know enough of some technical aspects of cyber-security to solve all the problems in this by myself in time, Tarlogic has promised to help (in a tutoring way) if a problem arises. This help could be critical to solve or get around some of the most complex problems, which probably happen to be critical points, needing to be dealt with to continue working on that stage.
Negative effects	Cascading delays.
Probability	Medium
Impact	Medium
Exposition	Medium
Indicator	A simple technical question takes more than 2 working days to be answered or a complex question takes more than 7 working days.
Prevention: Mitigate	Ask in a clear way and with as many details as possible. Ask during work hours, to ensure they are available.
Correction: Mitigate	Redo planning and possibly change the scope.

Identifier	<b>R-006</b>
Name	The learning curve of some technologies is larger than expected
Description	This is a critical need because not having enough knowledge can result in an inefficient approach to accomplishing the objectives.
Negative effects	Loss of quality. The work is more complicated.
Probability	Medium
Impact	Medium
Exposition	Medium
Indicator	The duration of the study of the technologies takes 50% than expected.
Correction: Mitigate	Redo planning and possibly change the scope. Ask the security consultants of Tarlogic for specific help. Maybe the need to rework completely some functionality.

Identifier	<b>R-007</b>
Name	The unexplained parts of the project take more time than expected
Description	There is not enough specification on what a task implies or not enough planning. This means that a part of the project is not understood as it should, and the work done is not what was expected or is not enough, needing more time to finish.
Negative effects	Wasted time that should have been easy to avoid. Loss of quality. Could mean the failure of the project if the objectives can not be accomplished in the time left.
Probability	Low
Impact	High
Exposition	Medium
Indicator	A task takes 25% more time than expected and when the causes are investigated it is revealed that there were ambiguous descriptions or planning.
Prevention: Avoid	Try to detail every part enough, having no obvious ambiguity.
Correction: Mitigate	Possible need to redo the specifications. Redo planning and possibly change the scope. Maybe having to redo related work.



Identifier	<b>R-008</b>
Name	Can not access source material
Description	All or part of the source material can not be accessed, probably because the only host of the resource is down.
Negative effects	In some cases this could mean a delay in a critical task, delaying the whole project for an unknown period of time.
Probability	Low
Impact	Medium
Exposition	Low
Indicator	There have been at least 10 failed attempts to download the source material, at least 5 with a computer A in a network X and at least 5 with a computer B in a network Y.
Prevention: Avoid	When possible choose the source with the best uptime.
Correction: Mitigate	Redo planning and possibly change the scope. Possible need to cut out the part of the project that depends on this source. Maybe find another source or wait to the original source to be accessible again.

Identifier	<b>R-009</b>
Name	Unexpected changes to any of the software used in the project
Description	<p>Changes to base software could affect this project directly or indirectly: programs could fail or not work as expected. This could mean any software changes, from simple syntax to API changes.</p> <p>In a project that does not work in a bleeding edge environment, like this, this occurrence should be very rare and even if it were to happen it would have to interfere with the part of the software this project uses, which (as this is not bleeding edge) normally would be backwards compatible.</p>
Negative effects	Minor delays.
Probability	Low
Impact	Low
Exposition	Low
Indicator	The software is not working as expected due to a change in another software version.
Prevention: Mitigate	When possible use software that follow good design guidelines and try to be backwards compatible.
Correction: Mitigate	Need to adapt the software to work as expected or remove the related functionalities.

Identifier	<b>R-010</b>
Name	Loss of work
Description	Due to a bad configuration management or something else, there is a loss of work related to this project.
Negative effects	<p>Need to do again the work already done but lost.</p> <p>Depending of the time needed to recover the work, there could be minor or very big delays, planning, changes to the scope of the project and even its failure.</p>
Probability	Low
Impact	High
Exposition	Medium
Indicator	The need to replicate already done work is greater than 30 minutes.
Prevention: Mitigate	<p>Take snapshots of key status for each virtual machine.</p> <p>Automate backing up the data and store the copies both in a cloud storage service and in a local disk.</p>
Correction: Mitigate	<p>Recover the last backup available of the work.</p> <p>If needed work even outside schedule and in holidays.</p>

Identifier	<b>R-011</b>
Name	Wrong management of the project's configuration
Description	The project's configuration is inefficient or lacks work. For example due to unclear changes or taking too long to commit changes.
Negative effects	Maybe the failure of the project if the objectives can not be accomplished in the time left. Possibly wrong baselines or identification of the configuration elements. It could be that it takes more time than expected to manage the project. The project suffer delays because the need to redo management work and/or planned tasks.
Probability	Medium
Impact	High
Exposition	High
Indicator	There are 3 delays because of the configuration of the project.
Prevention: Avoid	The configuration of the project should be just complex enough (without ambiguity, to ensure a proper management), but not too much complex (which would be hard to follow). Use of familiar and standard tools, like Git. Optionally use an easier to manage lifecycle. Study of the configuration management done in previous final degree projects, to get a proper idea of its scope and details.

Identifier	<b>R-012</b>
Name	A delay in one task leads to cascading delays in the dependent tasks
Description	A task gets delayed and one or more tasks depends on its completion to start, so they get delayed too.
Negative effects	Cascading delays.
Probability	Medium
Impact	Medium
Exposition	Medium
Indicator	At least 2 tasks are delayed, due to only one of them needing more time.
Prevention: Avoid	When planning, avoid task dependencies whenever possible. Optionally use a lifecycle based on increments.
Correction: Mitigate	Redo planning and possibly change the scope.

Identifier	<b>R-013</b>
Name	The student can not find a way to code the detection of a certain occurrence
Description	It could be that the knowledge of the student is too limited or the problem has too much logical or mathematical difficulty. Another possibility is that the event is impossible to detect with the current technologies. If so, this impossibility could be hard to assure too, due to the complexity of nowadays technology.
Negative effects	High difficulty to estimate the time needed to detect the event. Cascading delays.
Probability	Low
Impact	Low
Exposition	Low
Indicator	Finding a way to detect the occurrence takes 30% more time than planned.
Prevention: Mitigate	Have as much information on the problem as possible, the more detailed the better.
Correction: Mitigate	Ask the security consultants of Tarlogic for help. Demonstrate that it is possible to detect it.

Identifier	<b>R-014</b>
Name	The quality of the product is not enough
Description	The final result is does not comply the quality standard set for this project.
Negative effects	The incorporation to the official repository gets rejected. Redo planning and possibly change the scope. Analysis of the changes needed to improve the quality.
Probability	Low
Impact	High
Exposition	Medium
Indicator	Getting 10 suggestions to rework functionality.
Prevention: Avoid	Follow design patterns. Follow the design guidelines of the official repository when possible.
Correction: Mitigate	Need to redo and test work. Pass some kind of quality control.

Identifier	<b>R-015</b>
Name	Sickness or overwork
Description	The health of the student deteriorates to the point it affects the project.
Negative effects	Probably the quality of the project drops. Possibly delays, that could be hard to specify their limit. Analysis of the changes needed to improve the quality. In the worst case scenario the project can not continue and fails.
Probability	Medium
Impact	High
Exposition	Medium
Indicator	There is an unexpected delay because the functionality is not done but there has not been any important issues that could explain it but there is a clear deterioration of the student health.
Prevention: Avoid	Stay healthy by following a regular schedule for work and exercising, that includes multiple rest periods. Optionally maintain a diet.
Correction: Mitigate	Go to the doctor and follow any instructions to improve the recovery.

Identifier	<b>R-016</b>
Name	Performance issues
Description	The program is too heavy for the environment and takes too much resources, because there are not good enough optimizations or the problems are poorly approached.
Negative effects	Minor delays.
Probability	Low
Impact	Low
Exposition	Low
Indicator	The program takes 30% more resources that at the beginning of the project.
Prevention: Mitigate	If possible use efficient algorithms and check the efficiency after the testing is done for each increment.
Correction: Mitigate	Analysis of faster ways to solve the problem. Code and test a faster solution.

Identifier	<b>R-017</b>
Name	Unnecessary work
Description	Resources are wasted in work that latter is not used. This could happen because multiple reasons, like wrong assumptions or balancing of the remaining time of the project.
Negative effects	Minor delays.
Probability	Low
Impact	Low
Exposition	Low
Indicator	There is at least one functionality not necessary or useful for any requirement.
Prevention: Avoid	In the design stage make sure that everything is really needed.
Correction: Mitigate	Evaluate again if the work planned is really needed.

Identifier	<b>R-018</b>
Name	Optional requirements delay the project
Description	Optional requirements get too much time or are treated as vital.
Negative effects	The task related to these requirements get too much resources. Vital requirements get less resources, making the project loss value.
Probability	Low
Impact	Low
Exposition	Low
Indicator	There is at least one functionality from an optional requirement, when the project is behind schedule and there are vital requirements not yet accomplished.
Prevention: Avoid	The optional requirements are planned as optional: they are only done if there is enough time left.
Correction: Mitigate	Redo the planning.

### 2.4.5 Risk supervision