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

List of Figures

List of Tables

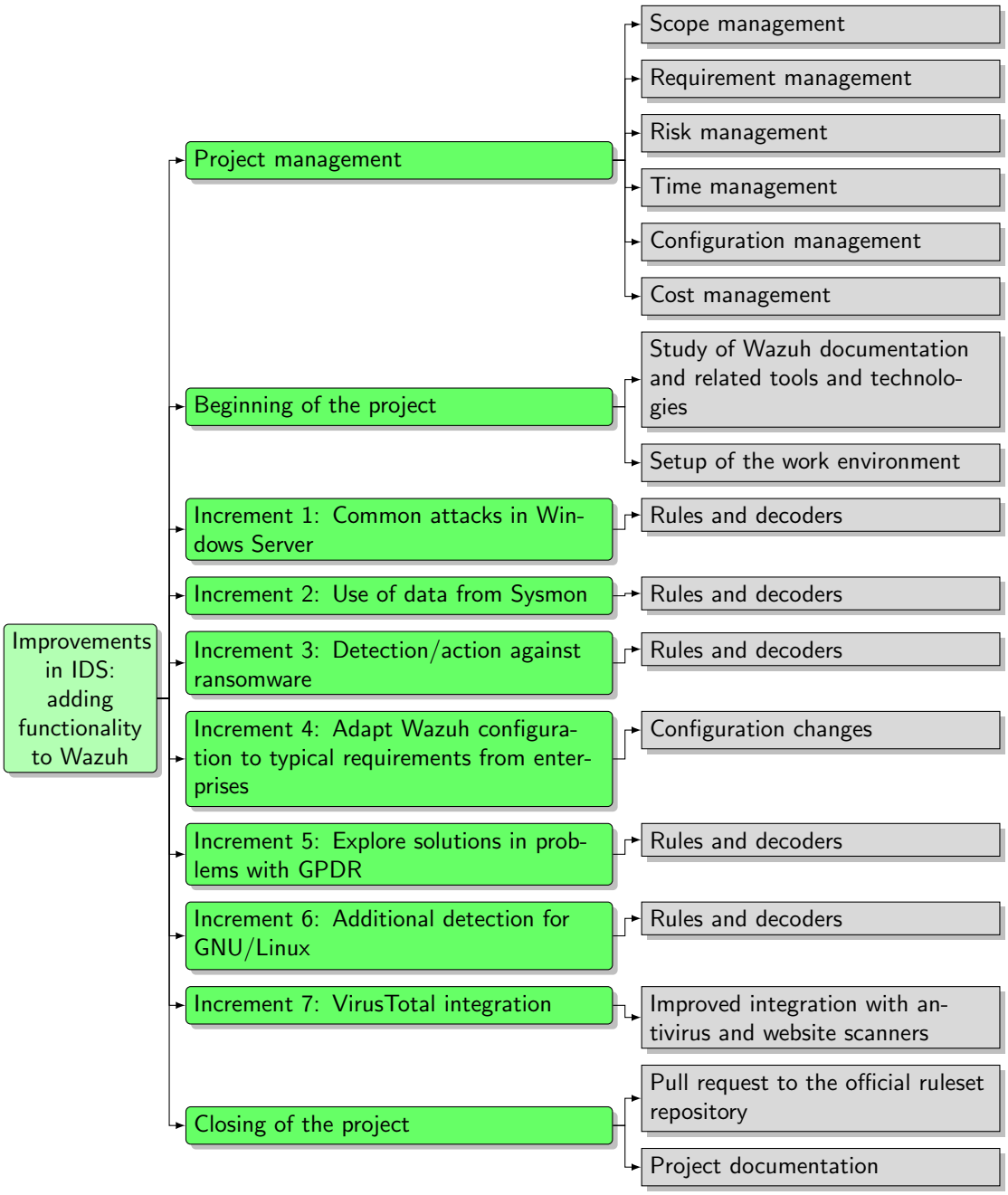
Chapter 1

Introdución

Chapter 2

Planning

2.1 Initial WBS



2.2 Initial planning

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

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.

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.

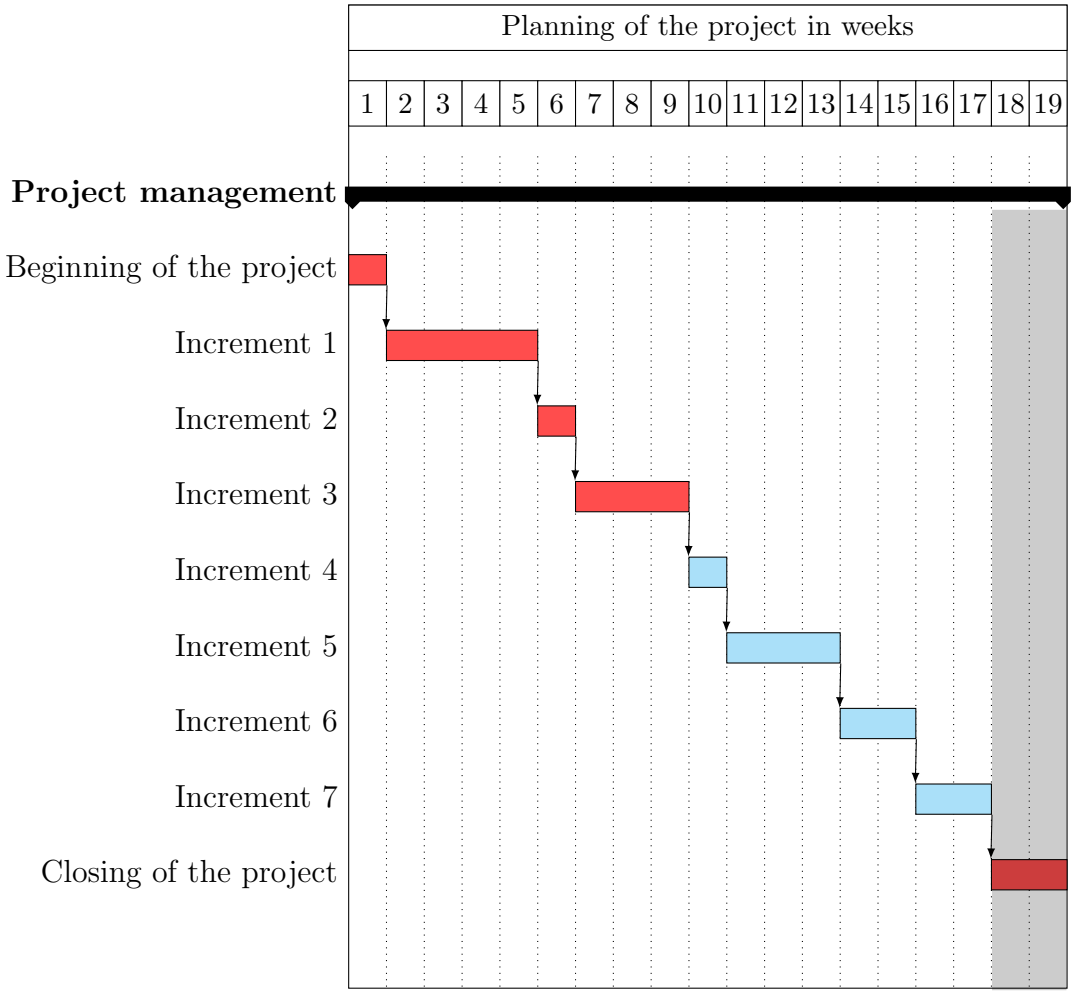


Figure 2.1: Initial planning

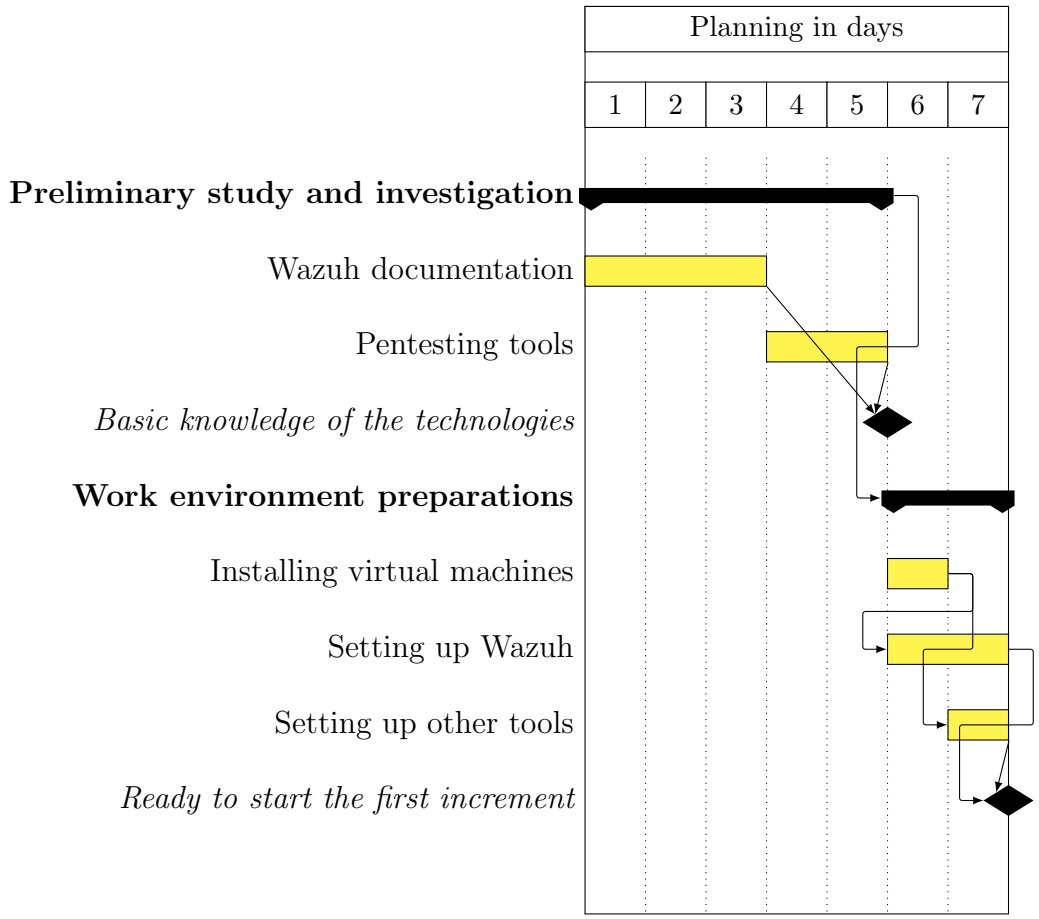


Figure 2.2: "Beginning of the project" planning

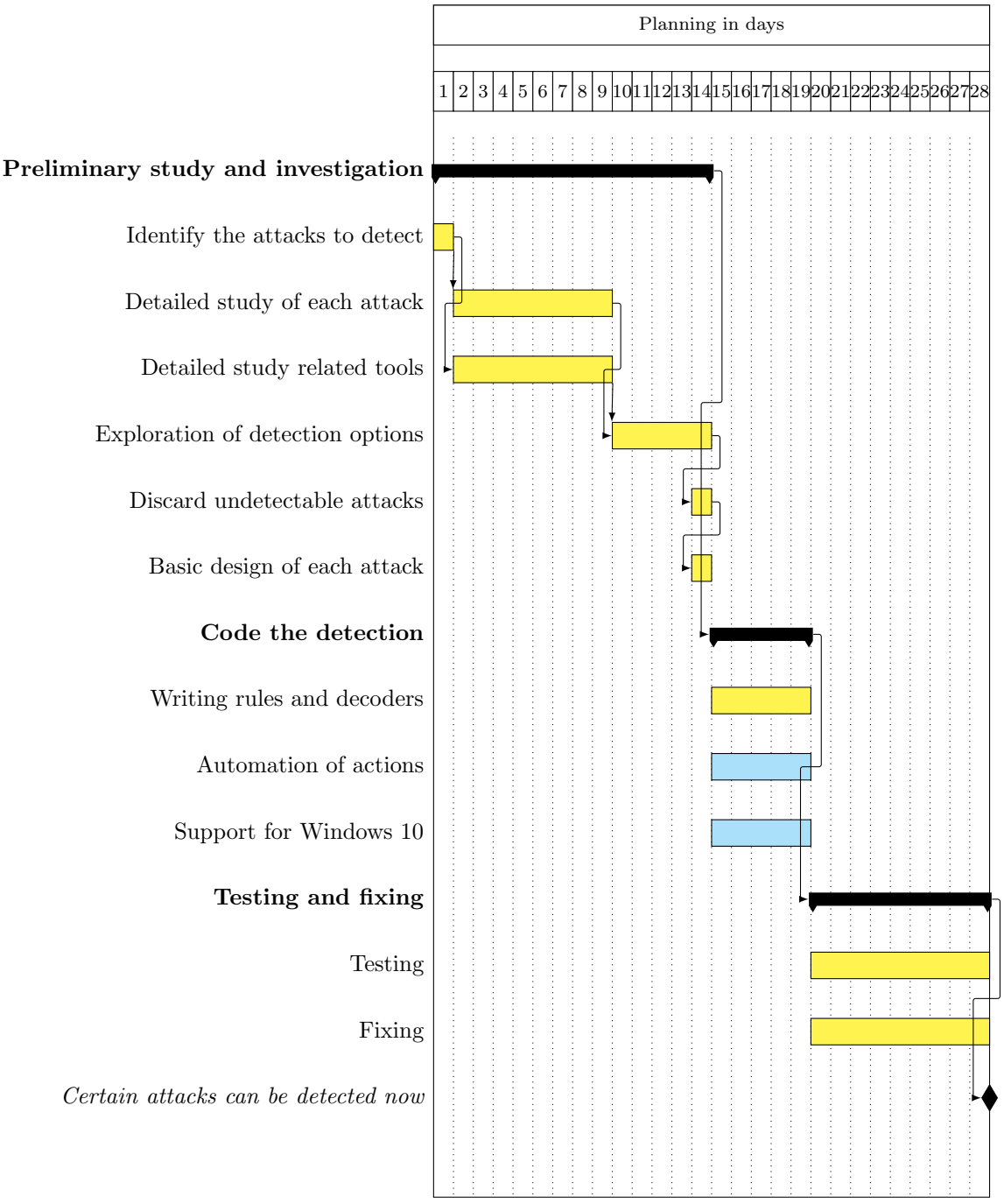


Figure 2.3: "Increment 1: Common attacks in Windows Server" planning

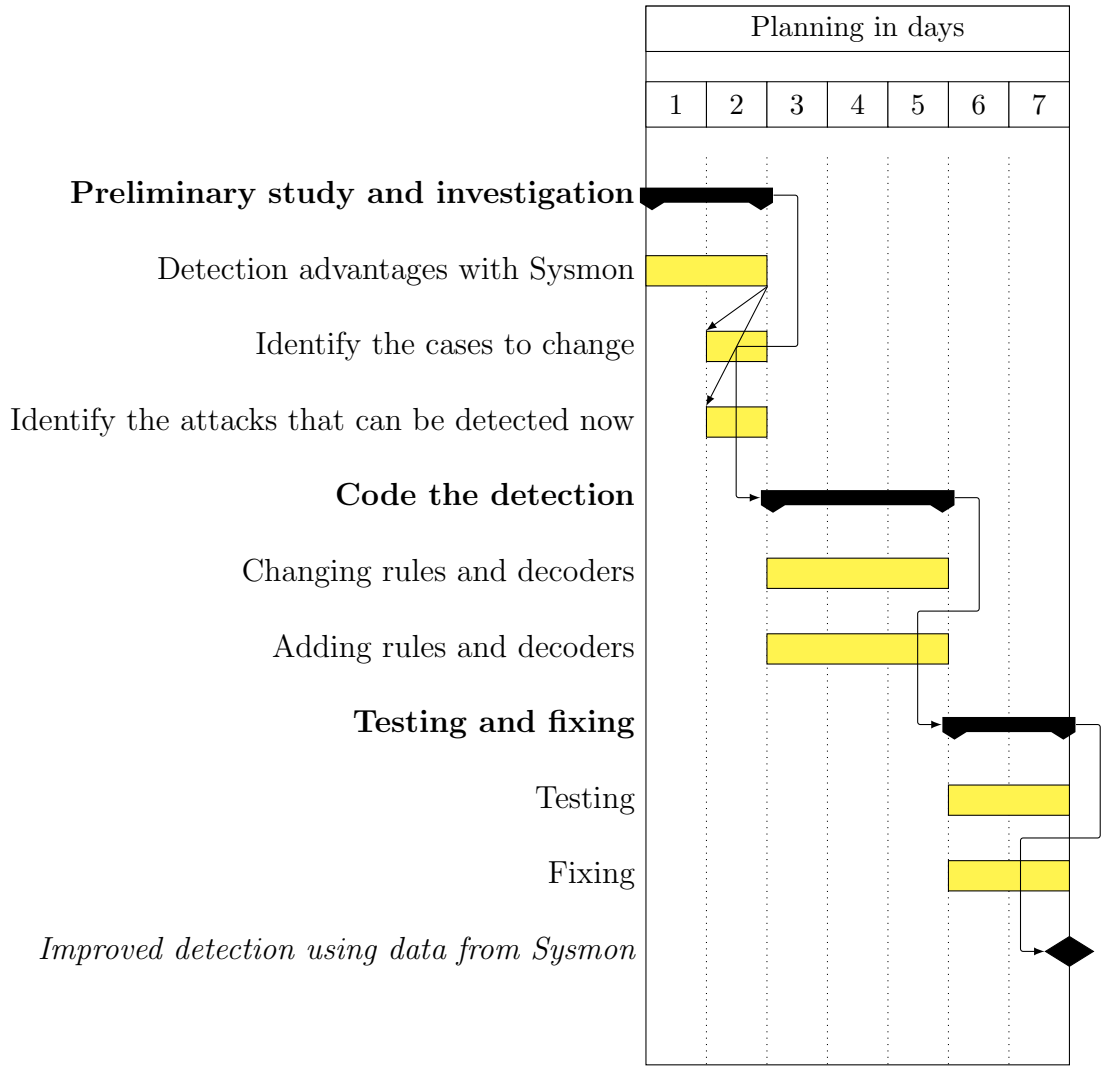


Figure 2.4: "Increment 2: Use of data from Sysmon" planning

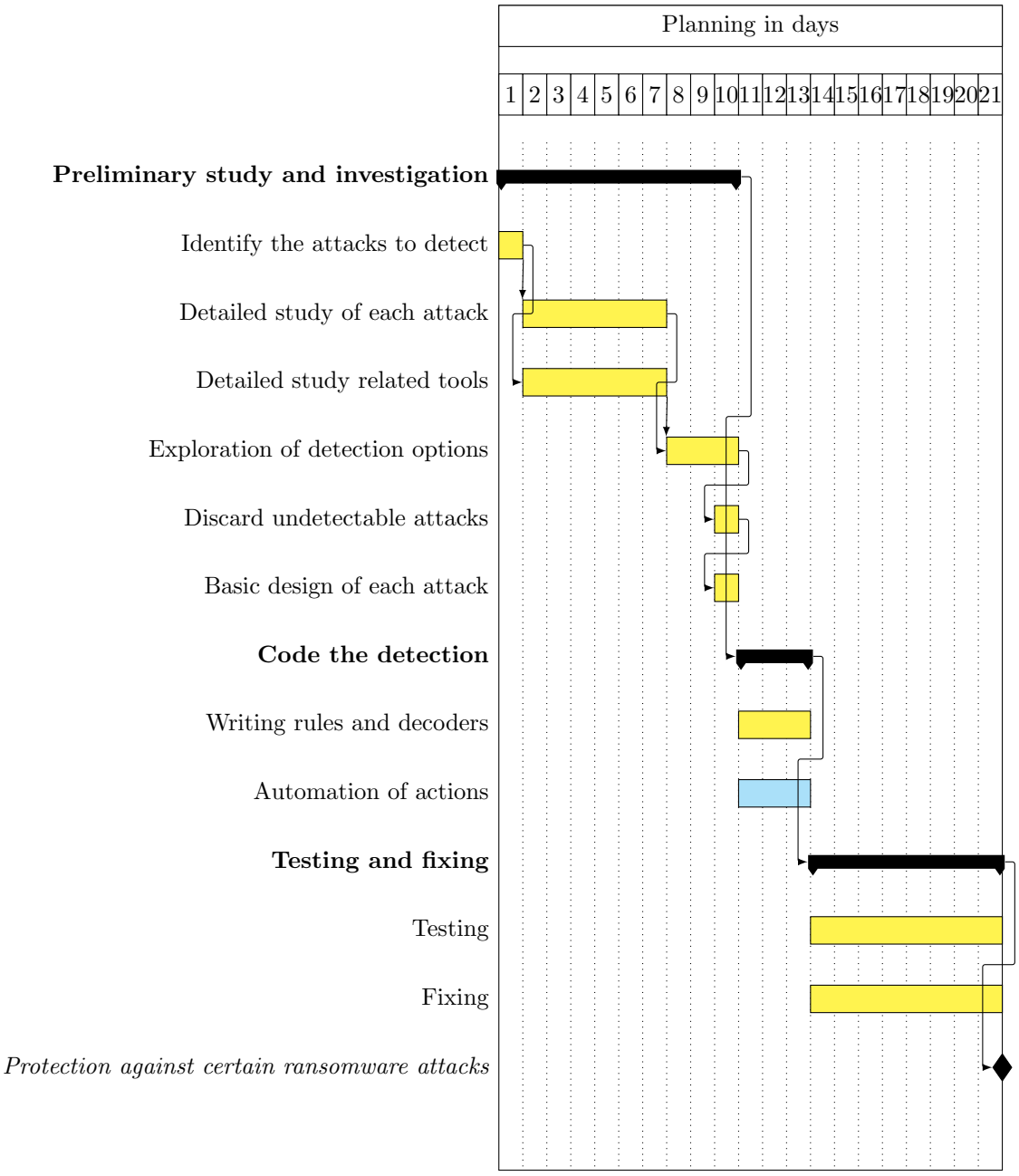


Figure 2.5: "Increment 3: Detection/action against ransomware" planning

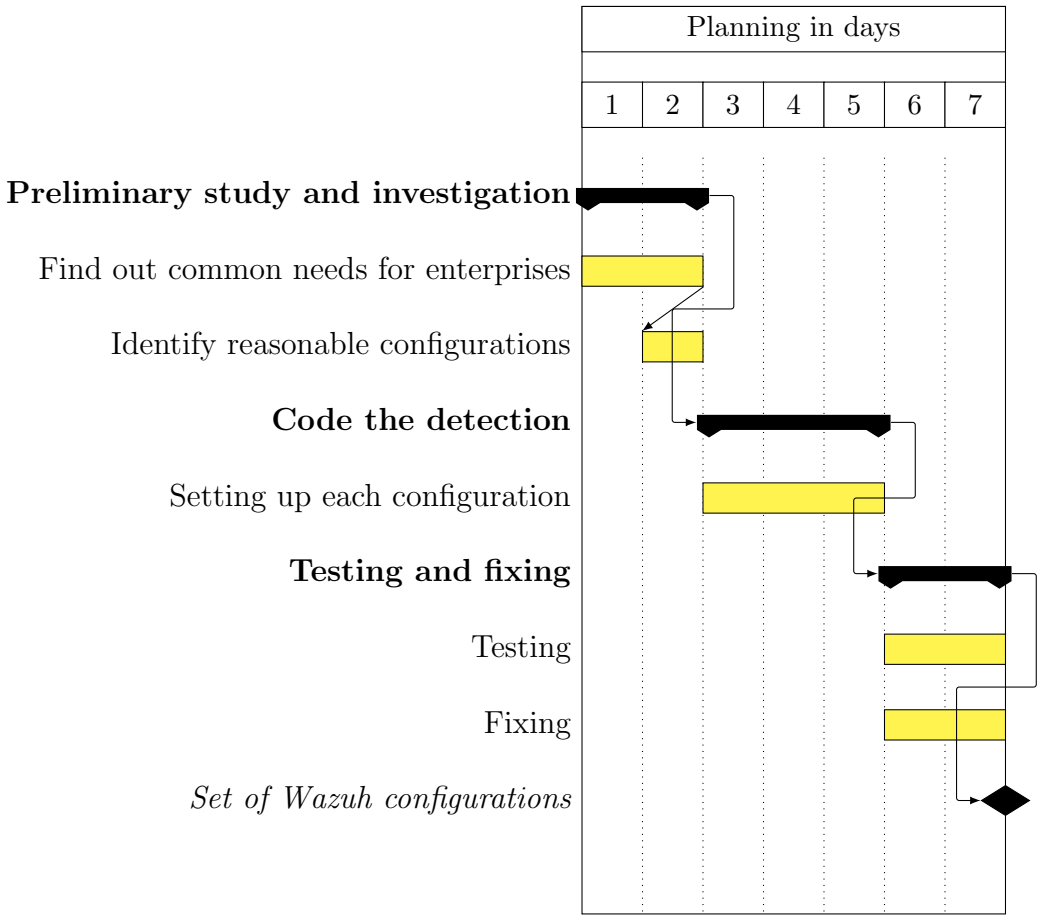


Figure 2.6: "Increment 4: Adapt Wazuh configuration to typical requirements from enterprises" planning

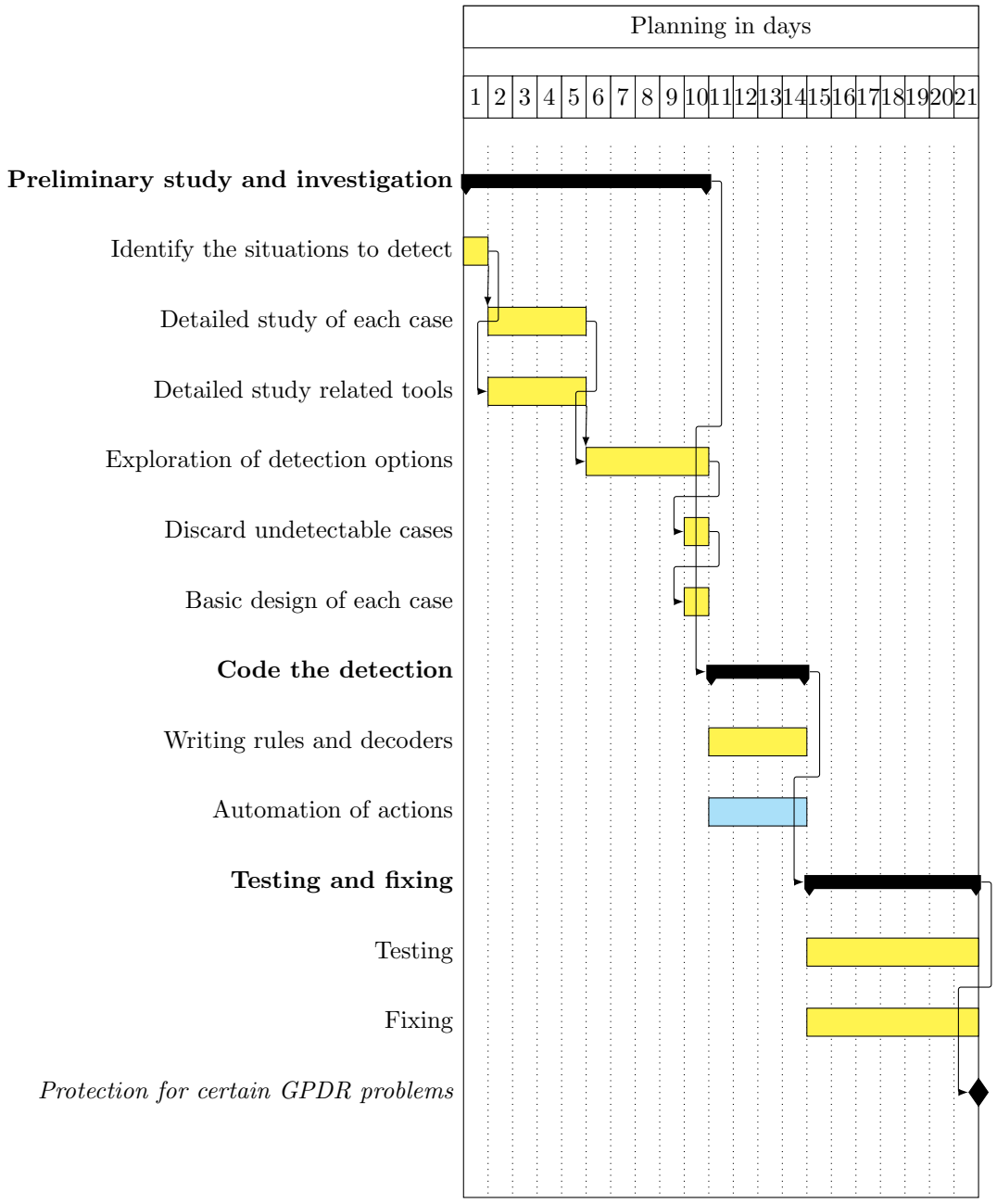


Figure 2.7: "Increment 5: Explore solutions in problems with GPDR" planning

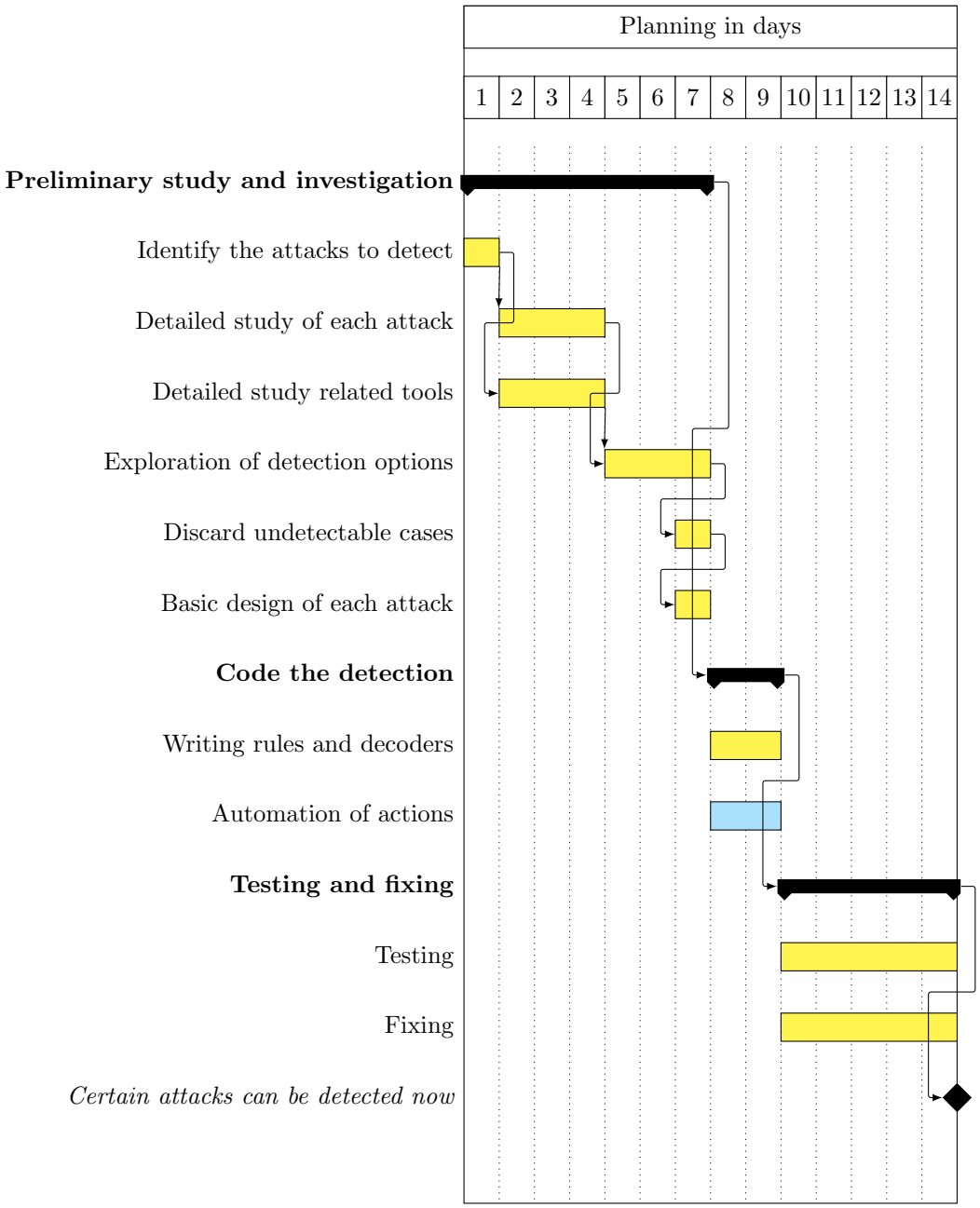


Figure 2.8: "Increment 6: Additional detection for GNU/Linux" planning

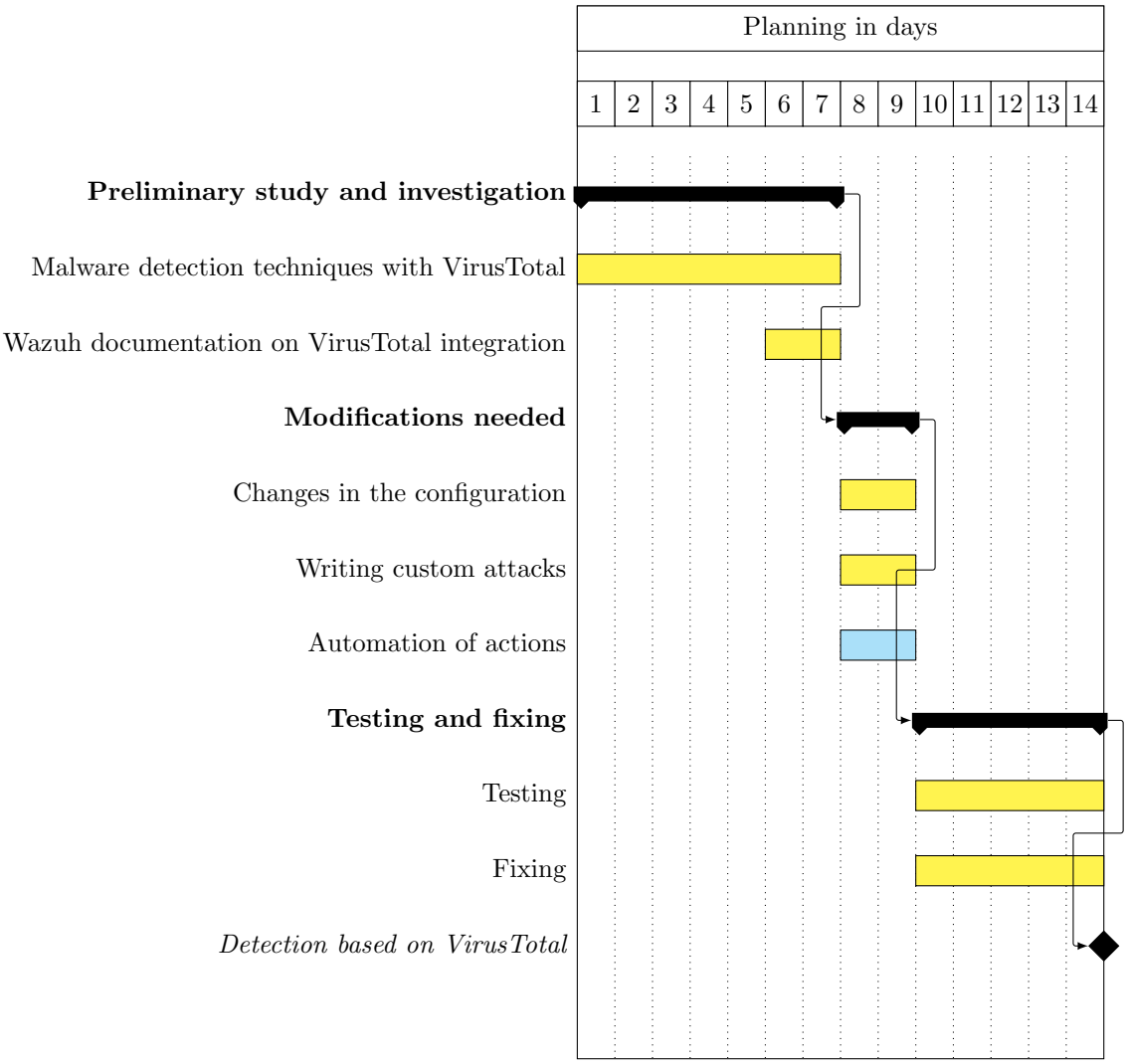


Figure 2.9: "Increment 7: VirusTotal integration" planning

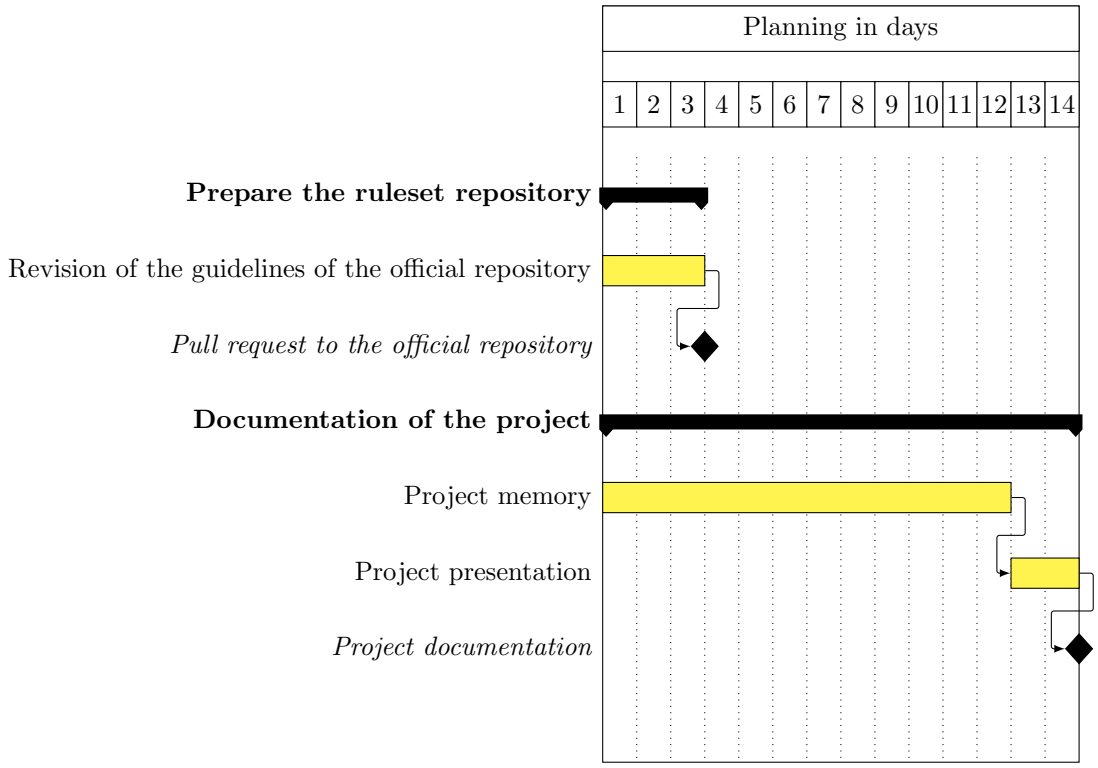


Figure 2.10: "Closing of the project" planning

2.3 Final planning

Chapter 3

Requirements

Chapter 4

Design

Chapter 5

Conclusions and additions

5.1 Risk management

5.1.1 Risk metrics

Chances of the risk happening	Probability
≥80%	High
Between 30% and 80%	Medium
≤30%	Low

Table 5.1: Probability classification of risks

Resource in Place / Effort / Cost	Impact
≥20%	High
Between 10% and 20%	Medium
≤10%	Low

Table 5.2: Impact classification of risks

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

Table 5.3: Method of calculation of Exposition based of Probability and Impact

5.1.2 Risk types

5.1.3 Risk identification

Table 5.4: Project risks

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 detect 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 increment the time need to complet the project

5.1.4 Risk analysis

Identifier	**R-001**
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. Rework the planning. Cascading delays.
Probability	Medium
Impact	High
Exposition	High

Identifier	**R-002**
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	Need to redo the analysis of specifications. Redo planning. Rework of related requirements and work based on them, including the need to test the results. Possible failure of the project if the objectives can not be accomplished in the time left.
Probability	High
Impact	High
Exposition	High

Identifier	**R-003**
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

Identifier	**R-004**
Name	Lack of key information from sources
Description	Not having key information from articles, documentation or manuals.
Negative effects	Minor delays. Added difficulty, increasing the resources needed. Need to rework and test the functionality, even completely, to follow the desired procedure.
Probability	Medium
Impact	Medium
Exposition	Medium

Identifier	**R-005**
Name	Lack of feedback or support from the security consultants of Tarlogic
Description	Because I do not know enough of some technical aspects of cibersecurity 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

Identifier	**R-006**
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	The work is more complicated.
Probability	Medium
Impact	Medium
Exposition	Medium

Identifier	**R-007**
Name	The unexplained parts of the project take more time than expected
Description	There is not enough specification on what a tasks 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	Could mean the failure of the project if the objectives can not be accomplished in the time left.
Probability	Low
Impact	High
Exposition	Medium

Identifier	**R-008**
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

Identifier	**R-009**
Name	Unexpected changes to any of the software used in the project
Description	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. 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.
Negative effects	Minor delays.
Probability	Low
Impact	Low
Exposition	Low

Identifier	**R-010**
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	Need to do again the work already done but lost. 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.
Probability	Low
Impact	High
Exposition	Medium

Identifier	**R-011**
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	Wrong baselines or identification of the configuration elements. It takes more time than expected to manage the project. Maybe the failure of the project if the objectives can not be accomplished in the time left. This means the project suffer delays because the need to redo management work and/or planned tasks.
Probability	Medium
Impact	High
Exposition	High

Identifier	**R-012**
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

Identifier	**R-013**
Name	The student can not find a way to detect 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. It could be that there is impossible to detect the event with the current technologies, if so this impossibility could be hard to assure, due to the complexity of now a days technology.
Negative effects	Cascading delays.
Probability	Low
Impact	Low
Exposition	Low

Identifier	**R-014**
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

Identifier	**R-015**
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

Identifier	**R-016**
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. Analysis of faster ways to solve the problem. The need to code and test a faster solution.
Probability	Low
Impact	Low
Exposition	Low

Identifier	**R-017**
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

Identifier	**R-018**
Name	Optional requirements increment the time need to complet 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

5.1.5 Risk planning

Identifier	**R-001**
Name	Optimist planning, “best case” (instead of a realistic “expected case”)
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	Reduce the scope of the project, leaving out initially planned increments.

Identifier	**R-002**
Name	Bad requirement specification
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	Reduce the scope of the project.

Identifier	**R-003**
Name	Design errors
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	**R-004**
Name	Lack of key information from sources
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 assistance. Maybe the need to rework completely some functionality.

Identifier	**R-005**
Name	Lack of feedback or support from the security consultants of Tarlogic
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.
Correction: Mitigate	Redo planning and possibly change the scope.

Identifier	**R-006**
Name	The learning curve of some technologies is larger than expected
Indicator	The duration of the study of the technologies takes 50% than expected.
Correction: Mitigate	Redo planning and possibly change the scope. Maybe the need to rework completely some functionality.

Identifier	**R-007**
Name	The unexplained parts of the project take more time than expected
Indicator	A task takes 15% 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	**R-008**
Name	Can not access source material
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	**R-009**
Name	Unexpected changes to any of the software used in the project
Indicator	There are 3 failures due to a change in 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	**R-010**
Name	Loss of work
Indicator	The need to replicate already done work is greater than 30 minutes.
Prevention: Mitigate	Take snapshots of key status for each virtual machine. Automate backing up the data and store the copies both in a cloud storage service and in a local disk.
Correction: Mitigate	Recover the last backup available of the work. If needed work even outside schedule and in holidays.

Identifier	**R-011**
Name	Wrong management of the project's configuration
Indicator	There are 3 delays because of the configuration of the project.
Prevention: Avoid	The configuration of the project should be just complex enough (whithout 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. Study of the configuration management done in previous final degree projects, to get a proper idea of its scope and details.

Identifier	**R-012**
Name	A delay in one task leads to cascading delays in the dependent tasks
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	**R-013**
Name	The student can not find a way to detect a certain occurrence
Indicator	Writing code that detects the occurrence takes 30% more time than planned.
Prevention: Mitigate	Have as much information on the problem as possible.
Correction: Mitigate	Ask the security consultants of Tarlogic for help. Demonstrate that it is possible to detect it.

Identifier	**R-014**
Name	The quality of the product is not enough
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. Optionally pass some kind of quality control.

Identifier	**R-015**
Name	Sickness or overwork
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	**R-016**
Name	Performance issues
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.

Identifier	**R-017**
Name	Unnecessary work
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	**R-018**
Name	Optional requirements increment the time need to complet the project
Indicator	There is at least one functionality from an optional requirement, when the project is behind its schedule and there are vital requirements not yet accomplished.
Prevention: Avoid	The planning leaves the non-vital requirements for the end of the project.
Correction: Mitigate	Redo the planning.

5.1.6 Risk supervision

Appendix A

Manuais técnicos

Manuais técnicos: en función do tipo de Traballo e metodoloxía empregada, o contido poderase dividir en varios documentos. En todo caso, neles incluirase toda a información precisa para aquelas persoas que se vaian a encargar do desenvolvemento e/ou modificación do Sistema (por exemplo código fonte, recursos necesarios, operacións necesarias para modificacións e probas, posibles problemas, etc.). O código fonte poderase entregar en soporte informático en formatos PDF ou postscript.

Appendix B

Manuais de usuario

Manuais de usuario: incluirán toda a información precisa para aquelas persoas que utilicen o Sistema: instalación, utilización, configuración, mensaxes de erro, etc. A documentación do usuario debe ser autocontida, é dicir, para o seu entendemento o usuario final non debe precisar da lectura de outro manual técnico.

Appendix C

Licenza

Se se quere pór unha licenza (GNU GPL, Creative Commons, etc), o texto da licenza vai aquí.