For use by the Project lecturer	Approved		Revision required	✗

Feedback

For use by the Project module lecturer only

Not clearly delineated.

Clearly describe how interaction between the embedded device that contains the WAF and the PC will work. It has sit between the PC and the network via ethernet cable?

Make explicit where machine learning fits in.

Revision required
(See Project Clickup site for submission deadline)

Symbol awarded: C

To be completed by the student				
PROJECT PROPOSAL 2024	Project no	TG4	Revision no	0
Title Mr	Surname **Becker**	Initials **JM**	Student no 20426471	Study leader (title, initials, surname) Mr. T. Green
Project title (the title on the project concept note) **Intelligent Web Application Firewall using JSON Web Token Inspection**				

Language editor details Ms. L.R.F. Fernandes	Language editor signature
Student declaration I understand what plagiarism is and that I have to complete my project on my own.	**Study leader declaration** This is a clear and unambiguous description of what is required in this project. Approved for submission (~~Yes~~/No)
Student signature Becker	Study leader signature and date

1. Project description

What is the problem to be solved with your project? What is your project about? What does your system have to do?

Web applications are currently increasingly susceptible to malicious attacks from threat actors. Web application firewalls (WAFs) are systems that intercept and inspect bidirectional web traffic between the web application and web clients to process and block this traffic. Current WAFs protect against vulnerabilities by inspecting session cookies or the claims of JSON web tokens (JWTs). These WAFs do not authenticate the user providing the JWT with the claims itself, merely inspecting the claim to estimate if it is a valid claim and that no harmful information is contained within. This means that current WAFs are susceptible to attacks where JWT claims could be changed to alter a different user's information. This is called JWT authentication bypass and next-generation WAFs are needed to protect against emerging threats such as these JWT authentication bypass and business logic flaws for which existing WAFs only provide limited protection.

The next-generation WAFs will have to use anomaly detection through machine learning to identify and block malicious requests by inspecting JWT claims to prevent JWT authentication bypass and business logic flaws. The system will also have to provide typical WAF features like preventing cross-site forgery & scripting (XSS), SQL and OS injections, and file inclusions which current WAFs have high protection rates against, but they lack advanced features like blocking JWT authentication bypass and business logic flaws.

2. Technical challenges in this project

Describe the technical challenges that are *beyond* those encountered up to the end of third year and in other final year modules.

2.1 Primary *design* challenges

Which aspects of the design of the system do you expect to be the most challenging?

- The first challenge is to achieve a balance between the true positive (malicious traffic) and the true negative (intended traffic) filtering accuracy.
- The second challenge is to design an accurate WAF where vulnerabilities such as cross-site scripting & forgery, file inclusion, SQL and OS injection, JWT authentication bypasses, and business logic flaws need to be prevented by efficiently and accurately implementing a next-generation WAF using JWT claims.
- Lastly, the WAF needs to build a baseline for a web application which can then be applied as a profile to different web applications. The baseline can be trained per web application to provide security to a range of applications.

2.2 Primary *implementation* challenges

Which aspects of the implementation to you expect to be the most challenging?

- The use and integration of a range of web applications will be a major challenge to implement on limited hardware due to memory constraints.
- The use and integration of client testing software on an embedded device is expected to put strain on it.
- The integration of the WAF on an embedded device is expected to put a strain on some operating system resources.
- The integrated system is expected to drop traffic if the system is overwhelmed, and this needs to be taken into account so as not to leak memory.

3. Functional analysis

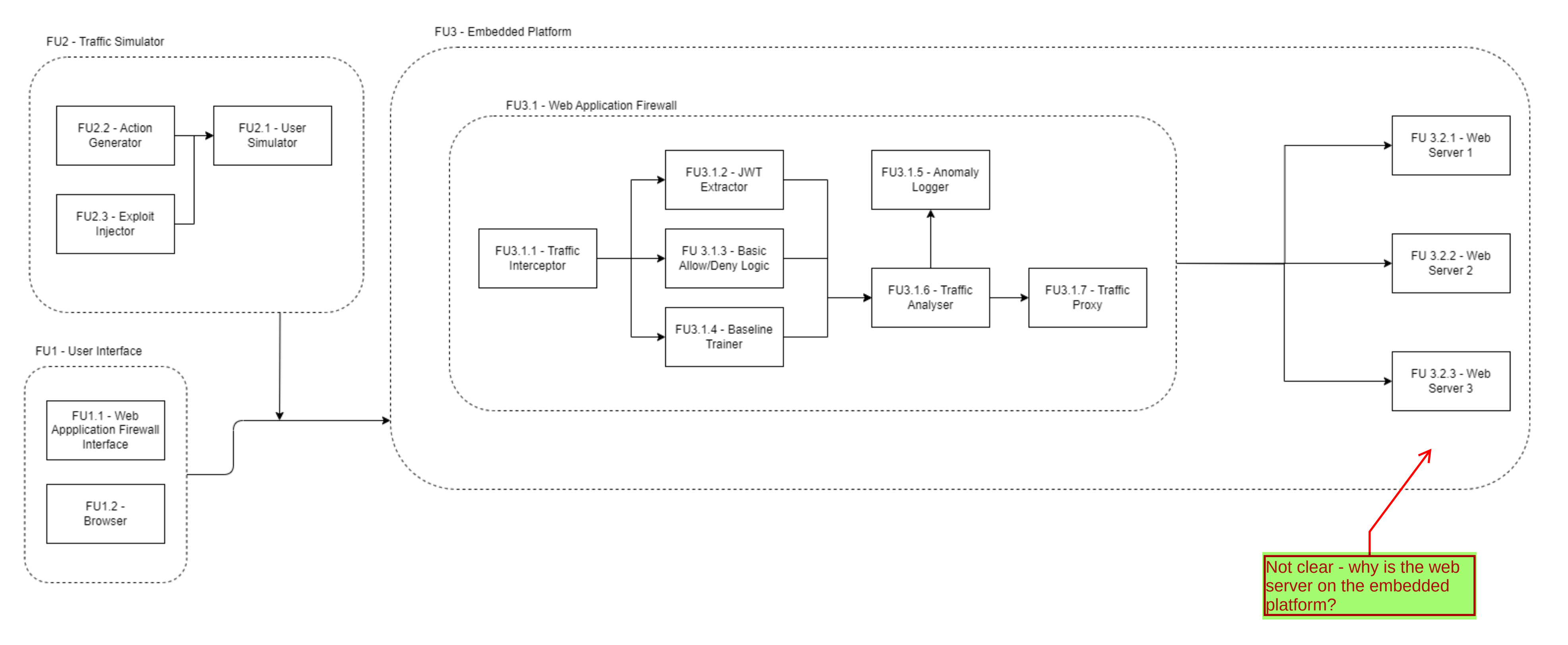
3.1 Functional description

Describe the design in terms of system functions as shown on the functional block diagram in section 3.2. This description should be in *narrative format*. **DO NOT** use a bullet list.

The next-generation WAF starts at the web client (FU 1), where the WAF administrator interacts with the WAF management interface (FU 1.1) and the web browser in which the user interacts with the web applications (FU 1.2). The traffic simulator (FU 2) then monitors the traffic and uses the action generator (FU 2.2) to simulate a specific web request. The exploit injector (FU 2.3) then injects the malicious payload into the request. From here, the request is passed to a user simulator to validate it and make it appear as if it comes from the user (FU 2.1). The request arrives at the WAF before entering the web application, where it is intercepted (FU 3.1.1) and simultaneously passed to the JWT extractor (FU 3.1.2), which decodes the JWT and extracts the claims. The basic allow/deny logic unit (FU 3.1.3) performs some simple pre-processing on the payload, and the baseline trainer (FU 3.1.4) stores the payload/token and awaits the allow/deny result. From here, the traffic passes through the traffic analyzer (FU 3.1.5), which has all the claim information needed. The claims are processed by different functions, each of which assigns an anomaly score to the request. If the anomaly score exceeds the threat threshold, the request is denied. Otherwise, it is allowed. If the request is denied, it is logged in the threat logger (FU 3.1.6), and a 406 error is sent to the user. If it is allowed, it is passed to the traffic proxy (3.1.7), which sends the request to the correct web server (FU 3.2.1-3.2.3).

If the request is valid, the web server processes it and generates a response. This response is sent back through the WAF, where it is processed with the same scrutiny as the request.

3.2 Functional block diagram (this should not be a flow diagram)



Not clear - why is the web server on the embedded platform?

4. System requirements and specifications

These are the core requirements of the system or product (the mission-critical requirements) in table format **IN ORDER OF IMPORTANCE**. Requirement 1 is the most fundamental requirement.

	Requirement 1: the fundamental functional and performance requirement of your project	**Requirement 2** (Number 2 in the order of importance)	**Requirement 3** (Number 3 in the order of importance)
1. Core mission requirements of the system or product. Focus on requirements that are **core** to solving the engineering problem. These will reflect the solution to the problem.	The WAF that will be created must be able to perform as well as current WAFs in preventing conventional attacks and better at preventing unconventional attacks.	The created WAF must be able to prevent conventional attacks such as client and server-side injection attacks with the same precision as existing WAFs.	The created WAF must be able to prevent unconventional attacks, such as server-side authentication bypass and business logic flaw attacks, with better precision than existing WAFs.
2. What is the target specification (in *measurable* terms) to be met in order to achieve the requirement in 1. above?	The target specification is to achieve a 90% overall balanced filtering efficacy, i.e. the average of the percentage of intended traffic passed through and the percentage of malicious traffic blocked.	The percentage of conventional attacks that will be blocked is 95%.	The percentage of unconventional attacks that will be blocked is 75%.
3. Motivation: Defend the specific target specification, i.e. the value that you selected. I.e., *why* will meeting the specification given in point 2 above *solve the problem*?	The overall balanced filtering efficacy is chosen to be 90% as this is on par with current products in the industry, and this is because if 90% of flagged requests are blocked, an attacker cannot chain together attacks.	The target specification is 95% because conventional attacks are still the most common types, representing around 75%, and therefore, the efficacy should be 5% higher than the overall balanced filtering efficacy.	The target specification is 75% as unconventional attacks are still the least common types of attacks, around 25%, and should thus be 15% lower than the overall balanced filtering efficacy.
4. How will you demonstrate at the examination that this requirement and specification (points 1 and 2 above) have been met? Be explicit about how you will *prove* these were met.	A test suite will send a malicious or valid request to the web application where it will be intercepted by the WAF. The test suite will then provide a summary of results on the efficacy of the WAF for the requests sent.	A test suite will attempt a conventional malicious attack on the web application where it will be intercepted by the WAF. The test suite then provides a summary of the results of the WAF's efficacy towards conventional attack vectors.	The target specification is 75% as unconventional attacks are still the least common types of attacks, around 25%, and should thus be 15% lower than the overall balanced filtering efficacy.
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*	A test suite will be designed and implemented, as well as JWT decoding, WAF functions, and several web applications, each with a different vulnerability.	Conventional WAF functions such as IP fencing, geo-fencing and geo-blocking, request and response inspection, security rules, anomaly scoring, DDoS rate limits, and bot mitigation will be designed and implemented.	Unconventional WAF functions, such as anomaly detection, will be designed and implemented.
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none". Clearly specify for what tasks **library functions** will be used (if relevant to the project).	The embedded hardware platform and the operating system software will be taken off the shelf.	The aspects that will be taken off the shelf are the threat thresholds commonly used in industry.	None.

Make explicit where machine learning fits in

System requirements and specifications page 2

	Requirement 4	**Requirement 5**	**Requirement 6**
1. <u>Core mission requirements of the system or product.</u> Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.	The system must be able to classify and allow/deny traffic with minimal latency.	The WAF must be able to serve multiple different web applications simultaneously.	
2. What is the <u>target specification</u> (in measurable terms) to be met in order to achieve the requirement in 1. above?	The target specification is not to add more than a 50% delay to the original response time.	The target specification is to host 3 web applications that will be served in parallel by the WAF on one web server.	
3. Motivation: Defend the <u>specific</u> target specification, i.e. the <u>value that you selected</u>. I.e., *why* will meeting the specification given in point 2 above solve the problem?	The target specification is chosen since most response times are between 200 - 1000 ms, and 50% does not add much of a noticeable delay to the response time.	Generally, one processing core per web application is required, and since most modern microcontrollers have a quad-core processor, one core is for the WAF, and three are for the web application. This can be scaled on larger servers.	
4. How will you <u>demonstrate at the examination</u> that this requirement and specification (points 1 and 2 above) have been met? Be explicit about how you will *prove* these were met.	It will be demonstrated by disabling the WAF on the client and measuring the delay, then enabling the WAF and measuring the delay again for a large dataset to calculate and compare the average response time delay added.	Three different tabs on a web browser will be opened, each containing a different web application hosted on the web server with one destination address.	
5. <u>Your own design contribution:</u> what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? <u>If none,</u> <u>*remove this requirement.*</u>	The design and implementation of the time delay measurement function. Revise, a "delay measurement is not final year design. Of course the entire algorithm needs to be optimized to achieve the latency requirements	To decide which vulnerabilities each web server will have and the hostname and IP address for each web server. This is not design - revise or remove.	
6. What are the aspects to be <u>taken off the shelf</u> to meet this requirement? If none, indicate "none". Clearly specify for what tasks **library functions** will be used (if relevant to the project).	The function to retrieve the current system time will be taken off the shelf.	Existing vulnerable web applications will be taken off the shelf.	

System requirements and specifications page 3

	Requirement 7	Requirement 8	Requirement 9
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.			
2. What is the target specification (in measurable terms) to be met in order to achieve the requirement in 1. above?			
3. Motivation: Defend the specific target specification, i.e. the value that you selected. I.e., *why* will meeting the specification given in point 2 above solve the problem?			
4. **How will you demonstrate at the examination** that this requirement and specification (points 1 and 2 above) have been met? Be explicit about how you will *prove* these were met.			
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*			
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none". Clearly specify for what tasks **library functions** will be used (if relevant to the project).			

System requirements and specifications page 4

	Requirement 10	Requirement 11	Requirement 12
1. Core mission requirements of the system or product. Focus on requirements that are core to solving the engineering problem. These will reflect the solution to the problem.			
2. What is the target specification (in measurable terms) to be met in order to achieve the requirement in 1. above?			
3. **Motivation:** Defend the specific target specification selected, i.e. the value. *Why* will meeting the specification given in point 2 above *solve the problem*?			
4. **How will you demonstrate at the examination** that this requirement and specification (points 1 and 2 above) have been met? Be explicit about how you will *prove* these were met.			
5. Your own design contribution: what are the aspects that *you will design and implement yourself* to meet the requirement in point 2? If none, *remove this requirement.*			
6. What are the aspects to be taken off the shelf to meet this requirement? If none, indicate "none". Explicitly indicate what tasks library functions will be used for (if relevant to the project).			

5. Field conditions

These are the REAL WORLD CONDITIONS under which your project has to work and has to be demonstrated.

	Real world field condition 1	Real world field condition 2	Real world field condition 3
Field condition requirement. In which field conditions does the system have to operate? Describe the one, two or three most important field conditions.	The system must be able to operate in an enclosed and air-conditioned room at room temperature (24 degrees Celsius) to simulate data center conditions close to servers.		

6. Student tasks

6.1 Design and implementation tasks

List your primary design and implementation tasks in bullet list format (5-10 bullets). These are *not* product requirements, but *your* tasks.

- The web client, web server, and WAF need to be designed and implemented on an embedded platform.
- The system needs to be simulated on a PC using Python where hardware constraints are not a problem and then scaled to fit on the embedded platform.
- The system needs to effectively prevent SQL-Injection, Cross-Site Scripting, File Inclusion, and JWT authentication bypass attacks to protect the vulnerable web server.

6.2 New knowledge to be acquired

Describe what the theoretical foundation to the project is, and which new knowledge you will acquire (*beyond* that covered in any other undergraduate modules).

- The student needs to learn what JWTs are and how they are used to bypass authentication measures.
- The student needs to learn what WAFs are and how security vulnerabilities in web applications are prevented.
- The student needs to learn what classifier-based algorithms are and how they can be used to model user behavior and prevent out-of-the-ordinary behavior.
- The student needs to learn how web application servers and their clients communicate to intercept the communication and how that is exploited.
- The student needs to learn how packet filtering works and how it can be utilized in machine learning to predict user behavior.
- The student needs to learn what packet inspection is and how it is utilized by WAFs to determine the intent of the packet.