# Problem Statement

The problem being addressed is API rate limitation violations in Kubernetes-based microservice architecture, an increasing concern in cloud-native applications. As organisations increasingly employ APIs for communication across microservices, these APIs become subject to abuse, such as when attackers or misconfigured clients exceed the allowed number of requests per unit of time. This can lead to performance degradation, illegal access, or denial of service (DoS) attacks. Traditional centralised anomaly detection solutions, which aggregate data across services to uncover such abuses, face data protection and scalability issues, especially in distributed environments like Kubernetes.[3] Recent research shows that federated learning (FL), a decentralised machine learning technique, can overcome these difficulties by enabling each microservice to train anomaly detection models locally on API data without exchanging raw data. This promotes privacy and scalability while retaining model accuracy.

Evidence for the issue:

The issue of API rate-limiting violations is a growing security and performance concern in microservice architectures. El Malki et al. (2022) highlight the increasing prevalence of API abuse, specifically rate-limiting violations, and how they can severely degrade the performance and reliability of microservices. The research emphasises that, without proper detection and mitigation, API overuse can cause system downtime, reduce user satisfaction, and expose applications to potential security threats. Serbout et al. (2023) analyse machine learning-based anomaly detection methods specifically for detecting API rate-limiting violations. This research presents experimental evidence demonstrating the effectiveness of machine learning in identifying anomalous API behaviour in real-time, thereby enhancing both the security and efficiency of microservice-based systems. Collectively, these studies highlight that API rate-limiting violations are not only a theoretical concern but a real and present danger to microservice architectures.

Approach:  
To address the issue of detecting API rate-limiting violations in Kubernetes-based microservice architectures, this project will implement a combination of federated learning (FL) and pre-trained anomaly detection models:

* Federated Learning (FL): FL will be used to enable distributed training of anomaly detection models across microservices without sharing raw API data, preserving privacy. Each microservice will locally train an anomaly detection model based on API request patterns, particularly focusing on identifying rate-limiting violations. These local models will send updates to a central server for aggregation, ensuring a global view of the system’s API activity while keeping data private.
* Pre-Trained Anomaly Detection Models: Pre-trained models will be used to identify abnormal API usage patterns, such as violations of rate limits. By using pre-trained models, the time and complexity required for developing models from scratch will be reduced. These models will be fine-tuned in the federated learning process to detect anomalies related to API usage and rate-limiting violations.
* Kubernetes-based Microservice Simulation: The system will be implemented and tested in a Kubernetes cluster to simulate real-world microservice environments. Kubernetes’ ability to scale microservices makes it a platform for evaluating the effectiveness of federated learning in detecting API abuse in a dynamic, distributed setting.

# Intended Users

There is a critical need for this project because API abuse is becoming increasingly common as organisations move toward cloud-native architectures and microservices. Traditional centralised security mechanisms for detecting abuse are not scalable and present significant data privacy risks. As highlighted by El Malki et al. (2022), centralising API logs or anomaly detection data exposes the system to potential breaches, making it imperative to explore privacy-preserving techniques like federated learning. Moreover, Kubernetes-based microservices are highly dynamic, meaning that abuse like rate-limiting violations can quickly degrade performance and open potential attack vectors. Serbout et al. emphasise the effectiveness of machine learning models in detecting such violations but also point out the need for decentralised models to handle scalability and privacy in distributed environments. This project addresses those needs by demonstrating how federated learning can enhance security and privacy in API abuse detection as it aims to help the below users.

* **DevOps and Cloud Engineers:** These professionals manage Kubernetes-based microservice environments and require robust tools to monitor API usage, detect anomalies, and mitigate potential abuses. The findings from this will benefit them by providing a scalable, privacy-preserving solution for API monitoring.
* **Academic Researchers in Cybersecurity and Machine Learning:** Researchers studying federated learning, API security, and anomaly detection will benefit from the project’s contribution to understanding how federated learning can be applied in real-world distributed systems for detecting API abuse. This will add to the body of knowledge regarding privacy-preserving security mechanisms in microservices architectures.

# System Requirements, Project Deliverables, and Final Project Outcome

The functional to support federated learning, API rate-limiting violation detection, and Kubernetes-based microservice orchestration:

* **Federated Learning Framework**:
  + Tools - TensorFlow Federated and PySyft will be used to implement the federated learning setup.
* **Anomaly Detection**:
  + Pre-trained deep learning anomaly detection models (e.g., LSTMs, CNNs) that can detect abnormal API usage patterns indicative of rate-limiting violations.
* **API Monitoring**:
  + API Logs should include key metrics like API request count, time intervals, and response times for analysis.
  + Tools - Prometheus
* **Kubernetes Integration**:
  + The solution must be deployable in a **Kubernetes environment**.

The outcome of the project will be a working prototype of a federated learning-based system that can detect API rate-limiting violations in Kubernetes-based microservice architectures. The system will:

* Demonstrate the effectiveness of federated learning in a distributed microservice environment by ensuring that API data remains private while still enabling anomaly detection.
* Prove that pre-trained anomaly detection models can be successfully deployed to detect API rate-limiting violations in real time.
* Provide a performance comparison between federated learning-based detection and centralized anomaly detection, focusing on privacy, detection accuracy, and scalability.

# Literature Review

1. El Malki et al discuss API rate-limiting highlight the increasing threat posed by API abuse, particularly rate-limiting violations in microservice architectures. The study presents empirical data on how unchecked API abuse can cause service disruptions and security breaches. While the research provides insights into the prevalence of API abuse, it relies on centralized detection methods, which raise privacy concerns due to the centralization of sensitive API data​.
2. Serbout et al explores security vulnerabilities in Kubernetes-based microservices, noting that APIs are among the most vulnerable points in these architectures. The study examines various security solutions for microservices but points out that many rely on centralized data collection and analysis, raising scalability and privacy issues. The authors suggest that distributed security frameworks, such as FL, could offer a more scalable and privacy-friendly solution​.